Gouvernement du Canada
Ministère des Communications

Government of Canada
Department of Communications

Le Centre canadien de recherche sur l'informatisation du travail
Canadian Workplace Automation Research Centre

CA1
DTR290-
90
R26
v.1

41.D

Canada

CA1
DTR290-
90
R26
v.1

460724

41.D

Communications Canada
Canadian Workplace Automation Research Centre (CWARC)
Integrated Systems Directorate

TP 10575E

REPLACING A PSEUDO-PILOT BY SPEECH GENERATION AND RECOGNITION TECHNOLOGY

VOLUME I

Prepared for
The Transportation Development Centre
Policy and Coordination Group
Transport Canada

July 1990

APR 23 19

460724

DDN 460723

92-00976/

TP 10575E

Replacing a pseudo-pilot by speech generation and recognition technology

Volume I

by

Jean-Yves Fiset

École Polytechnique de Montréal
and
Canadian Workplace Automation
Research Centre

July 1990

NOTE: The opinions and views expressed in this report are those of the author, and do not necessarily reflect those of the Transportation Development Centre of Transport Canada.

Ce document est aussi disponible en français : "Remplacement d'un pseudo-pilote par des technologies de reconnaissance et de génération de la parole", TP 10575F.

Transport Canada / Transports Canada

PUBLICATION DATA FORM

1. Transport Canada Publication No.	2. Project No.	3. Recipient's Catalogue No.
TP 10575E	7366	

4. Title and Subtitle	5. Publication Date
Replacing a pseudo-pilot by speech generation and recognition technology, Volume I	July 1990
	6. Performing Organization Document No.
7. Author(s)	8. Transport Canada File No.
Jean-Yves Fiset	ZCD 1455-201-2
9. Performing Organization Name and Address Canadian Workplace Automation Research Centre 1575 Chomedey Blvd. Laval, Quebec H7V 2X2	10. DSS File No.
	11. DSS or Transport Canada Contract No.
12. Sponsoring Agency Name and Address Transportation Development Centre 200 René-Lévesque Blvd. West West Tower, Suite 601 Montreal, Quebec H2Z 1X4	13. Type of Publication and Period Covered Final
	14. Project Officer James D. Reid

15. Supplementary Notes (Funding programs, titles of related publications, etc.)

Volume II, Appendices, also available
French report "Remplacement d'un pseudo-pilote par des technologies de reconnaissance et de génération de la parole", TP 10575F, also available

16. Abstract

Objective : The objective of this report is to determine how to replace a human pseudo-pilot with a device using speech recognition and generation technology.

Method : First, the state of the applied research in North America and Europe is surveyed. Then, a custom language model for air traffic control is developed. Finally, the use of speech recognition and generation technology for this application is reviewed.

Results : Although several systems are in the development stage, none is in current use. The use of a frame-based language model is recommended in this application. Recommendations are made regarding the use of speech recognition and generation technology.

Conclusion : The development of a voice driven pseudo-pilot seems technically feasible. However, a choice must be made between a passive system (e.g. voice-driven simulator) and an active system (e.g. one that will coach the trainee). A choice must also be made between an entirely Canadian system and a system developed in cooperation with a partner already active in this field.

17. Key Words	18. Distribution Statement
Air traffic control, pseudo-pilot language, speech technology, artificial intelligence	Limited number of copies available from the Transportation Development Centre on request.

19. Security Classification (of this publication)	20. Security Classification (of this page)	21. Declassification (date)	22. No. of Pages	23. Price
Unclassified	Unclassified	–	xiv +105	–

TDC/CDT 79-005 Rev 88

Canada

Transports Canada Transport Canada

FORMULE DE DONNÉES POUR PUBLICATION

1. N° de la publication de Transports Canada TP 10575E	2. N° de l'étude 7366
3. N° de catalogue du destinataire	
4. Titre et sous-titre Replacing a pseudo-pilot by speech generation and recognition technology, Volume I	5. Date de la publication Juillet 1990
	6. N° du document de l'organisme exécutant
7. Auteur(s) Jean-Yves Fiset	8. N° de dossier — Transports Canada ZCD 1455-201-2
9. Nom et adresse de l'organisme exécutant Centre canadien de recherche sur l'informatisation du travail 1575, boul. Chomedey Laval (Québec) H7V 2X2	10 N° de dossier — ASC
	11. N° de contrat — ASC ou Transports Canada
12. Nom et adresse de l'organisme parrain Centre de développement des transports 200 ouest, boul. René-Lévesque Tour ouest, suite 601 Montréal (Québec) H2Z 1X4	13. Genre de publication et période visée Rapport final
	14. Agent de projet James D. Reid

15. Remarques additionnelles (Programmes de financement, titres de publications connexes, etc.)

Volume II : Annexes
Rapport en français " Remplacement d'un pseudo-pilote par des technologies de reconnaissance et de génération de la parole", TP 10575F, aussi disponible

16. Résumé

But : ce rapport vise à déterminer comment on peut remplacer un pseudo-pilote humain par un dispositif utilisant des techniques de reconnaissance et de génération de la parole.

Méthode : on examine d'abord l'état actuel de la recherche appliquée en Amérique du Nord et en Europe. On élabore ensuite un modèle sur mesure du langage des élèves contrôleurs. On examine alors l'utilisation des technologies de reconnaissance et de génération de la parole.

Résultats : on constate qu'il y a un certain nombre de systèmes en cours de développement, même si aucun n'est encore en exploitation. On propose la représentation par schémas comme modèle de langage pour un pseudo-pilote automatisé. On formule des recommandations sur l'exploitation des technologies de reconnaissance et de génération de la parole.

Conclusion : il semble techniquement faisable de réaliser un pseudo-pilote automatisé. On doit cependant choisir si on veut un système passif (i.e. un simulateur commandé par la voix) ou un système actif (i.e. qui conseillera l'élève-contrôleur). On doit aussi choisir si on veut un système réalisé entièrement au Canada ou si on désire établir une coopération avec un organisme qui développe actuellement des pseudo-pilotes automatisés.

17. Mots clés Contrôle aérien, pseudo-pilote, langage, technologies vocales, intelligence artificielle.	18. Diffusion Le Centre de développement des transports dispose d'un nombre limité de copies.

19. Classification de sécurité (de cette publication)	20. Classification de sécurité (de cette page)	21. Déclassification (date)	22. Nombre de pages	23. Prix
Non classifiée	Non classifiée	—	xiv + 105	—

CDT/TDC 78-005
Rev 88

Canada

ACKNOWLEDGEMENTS

The author wishes to thank the following people for their assistance during the different steps in this project:

James D. Reid, of the Transportation Development Centre, Policy and Coordination Group, Transport Canada, who gave us the chance to work on this project.

Raymond Descout, of the Canadian Workplace Automation Research Centre, Communications Canada, who provided us with an outstanding work environment and his expertise in the application of speech technology.

Jean-Marc Robert, of the École Polytechnique de Montréal, who encouraged us in our research and provided the academic environment for this project.

Françoise Néel, of the Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI), in Orsay, France, who kindly organized a tour of the LIMSI laboratory.

Michael J.A. Asbury, of the Civil Aviation Authority, Great Britain, who helped to organize a tour of various research centres in England.

It is impossible to mention all the others who contributed in different ways to this project, but their assistance was sincerely appreciated in all cases.

SUMMARY

This report presents the results of a study on the feasibility of replacing a pseudo-pilot for training air traffic controllers with a device using speech generation and recognition technology.

The first step in the project was an overview of applied research in North America and Europe. In Canada, ATS has developed a pseudo-pilot that can accept spoken commands in a radar control environment. Although no such system is yet in operation, ATS has produced a demonstrator. In the United States, the FAA is planning to evaluate two systems:

- The TOTS (Tower Operator Training System) by Logicon. This is a system initially developed for the U.S. Navy. It is to be sent to the FAA training centre in 1990, and to be operational in 1991. The system is used for visual flight exercises.

- The ATCoach system, by UFA, which uses artificial intelligence techniques complemented by voice entry, employed for radar control exercises.

In Europe, it appears that the two main players are France and Great Britain. In France, the Centre d'Études de la Navigation Aérienne (CENA) is co-operating with the Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI), attached to the Centre National de la Recherche Scientifique, and the Vecsys company. This has led to the design of a prototype pseudo-pilot, which appears to hold great promise in terms of speech recognition. In Great Britain, the Bournemouth Air Traffic Control Evaluation Unit plans to install a hybrid pseudo-pilot using both speech recognition and touch entry. The system is intended for use in more advanced exercises by experienced controllers, rather than for the training of novice controllers.

The second step in the project was the development of a custom-made model of the language used in simulation exercises in a Canadian context. It is based on the LIMSI model, and was designed using transcripts of exercises recorded at the

Transport Canada Training Institute in Cornwall. In this model, each message from a trainee controller is associated with a specific category. Each category represents a frame, which is one way of representing knowledge in artificial intelligence. This type of model tolerates some of the limitations of existing speech recognition systems, and simplifies the task of designing an intelligent automated pseudo-pilot. It is recommended that this type of model be used in an automated pseudo-pilot, by incorporating the appropriate processes for handling dialogue errors.

The last step in the study consisted in examining the use of speech generation and recognition technology, with specific reference to:

- the characteristics of the trainee controller's tasks;
- the characteristics of the technology available;
- ergonomic aspects.

Accordingly, we recommend the use of a digitized voice for the responses generated by the automated pseudo-pilot, along with the use of a speaker-dependent connected-word speech recognition system. It is also recommended that the voice samples necessary for the speech recognizer be obtained during the phraseology training step.

Finally, aside from the technical feasibility of an automated pseudo-pilot, we stress that it is important for the next steps in the project that the following questions be answered:

- What is an automated pseudo-pilot intended to do? It must be decided whether the pseudo-pilot is to play a passive role in the training of air traffic controllers (i.e., executing valid instructions), or whether it is to be active (i.e., advising the trainee controller in addition to executing valid instructions).
- Is it essential that the system be 100% Canadian? Depending on the answer to the first question (whether the pseudo-pilot is to be active or passive), it may be advantageous to arrange some sort of co-operation with organizations having substantial experience in this field.

TABLE OF CONTENTS

1.		INTRODUCTION	1
2.		THE CURRENT STATE OF APPLIED RESEARCH	3
	2.1	Canada	3
	2.2	United States	4
	2.3	Europe	9
	2.4	Summary of current applied research	13
3.		SELECTING AN AIR TRAFFIC CONTROL SITUATION FOR THE DESIGN OF A LANGUAGE MODEL	15
	3.1	Phraseology used in different air traffic control situations	15
	3.2	Description of the situation considered in this study	16
	3.3	The trainee controller's workstation	19
4.		SELECTING THE MODEL TO REPRESENT CONTROLLER-PILOT LANGUAGE	22
	4.1	Analysis of existing models	22
	4.2	Selecting the model to represent the language	27
5.		MODELLING THE OPERATIVE LANGUAGE	29
	5.1	Target dialogue	29
	5.2	The approach to developing the model	30
6.		FIRST STEP IN DEVELOPING THE MODEL FOR SIMPLE EXERCISES	34
	6.1	Examining the length of transmissions	34
	6.2	Obtaining the vocabulary	35
	6.3	Preparing preliminary frames	39
	6.4	Validating the use of frames in an exercise executed by different trainees	41
	6.5	Representing an aircraft	43
	6.6	Relations between the dialogue and the states of an aircraft	44
7.		SECOND STEP IN DEVELOPING THE MODEL FOR SIMPLE EXERCISES	47
	7.1	Rules for frame instantiation	47
	7.2	Handling dialogue errors	52

7.3 Calling successive frames 60
7.4 Resulting model 64

8. EXTENSION OF THE RESULTING MODEL 68
8.1 Adaptability of the model 68
8.2 Relaxing constraints on the concatenation of messages and the number of recipients of a message 74
8.3 Comments on the modelling of the language 78

9. REQUIREMENTS AND POSSIBILITIES FOR SPEECH RECOGNITION 79
9.1 Required speech recognizer features 79
9.2 Training with the speech recognizer 81
9.3 Effects of interaction with a machine on language behaviour 82

10. REQUIREMENTS AND POSSIBILITIES FOR SPEECH GENERATION 84
10.1 Pertinent features specific to the task 84
10.2 Description and characteristics of generation methods 85
10.3 Ergonomic aspects: perception of generated speech 89
10.4 Effects of artificially generated speech on language behaviour 94
10.5 Selecting a synthetic voice for the automated pseudo-pilot 95

11. CONCLUSION 98

REFERENCES 100

APPENDICES

Note : Appendices are located in Volume II of this report.

Appendix A: Excerpt from the course plan used at the Transport Canada Training Institute, Cornwall, Ontario

Appendix B: Transcripts of simple exercises

Appendix C: Vocabulary used by trainee controllers in simple exercises

Appendix D: Preparation of individual frames for simple exercises

Appendix E: Summary of frames and sub-frames for simple exercises

Appendix F: Assignment of ratings

Appendix G: Analysis of the error recovery process

Appendix H: Transcripts of complex exercises

Appendix I: Vocabulary used by trainee controllers in complex exercises

Appendix J: Draft of additional frames for complex exercises

Appendix K: Analysis of concatenated transmissions

Appendix L: Vocabulary used by pseudo-pilots

LIST OF FIGURES

Figure 1 Block-diagram of the ATCoach system 7

Figure 2 Block-diagram of the pseudo-pilot developed at LIMSI 11

Figure 3 Midtown regional control zone 17

Figure 4 Trainee controller's workstation 19

Figure 5 Language representation in tree-structure form 23

Figure 6 Language representation in frame form 25

Figure 7 First step in developing the language model 31

Figure 8 Second step in developing the language model 33

Figure 9 Distribution of sentence length for simple exercises 35

Figure 10 Relationship between the trainee-pilot dialog and the states of an aircraft 45

Figure 11 Skeleton diagram of the dialogue management system 65

Figure 12 Example of completed frame 66

Figure 13 Distribution of sentence length for complex exercises 69

Figure 14 Comparison of sentence length distribution for simple and complex exercises 70

Figure 15 Steps in the concatenation of words or sentence fragments 86

Figure 16 Main components of a text-to-speech synthesis system 88

LIST OF TABLES

Table I	Performance of the SSI speech recognizer for the ATCoach system	8
Table II	Results of the evaluation of the vocabulary size	36
Table III	Evaluation of the number of extrinsic elements	38
Table IV	Size of operative vocabulary	39
Table V	Example of the development of a frame	40
Table VI	Representation of frame HDR	41
Table VII	Proportion of messages described by the various frames	42
Table VIII	Representation of the various states of an aircraft	43
Table IX	Relations between the representation of an aircraft and frames	49
Table X	Limits for the instantiation of various sub-frames	51
Table XI	Classification of dialog errors	53
Table XII	Error-handling methods for errors with self-correction	55
Table XIII	Method of handling requests for repetition issued by the trainee controller	56
Table XIV	Method of handling requests for repetition issued by the pseudo-pilot	57
Table XV	Methods for identifying and resolving ambiguities	58

Table XVI	Calling successive frames	62
Table XVII	Size of the incomplete net operative vocabulary for terminal control exercises	71
Table XVIII	Size of the operative vocabulary for terminal control exercises	71
Table XIX	Usefulness of simple model frames for terminal control situations	73
Table XX	Analysis of transmissions containing several messages for one recipient	75
Table XXI	Analysis of transmissions containing different messages for more than one recipient	77
Table XXII	Vocabulary used by pseudo-pilots in the four transcripts	85
Table XXIII	Summary of results of clarity tests, for text-to-speech synthesis systems	91

1. INTRODUCTION

This study, conducted under a research contract awarded by Transport Canada, is intended to establish how a human pseudo-pilot can be replaced with an automated device using speech recognition and generation. A pseudo-pilot is an individual that operates a simulator, following the oral instructions given by a trainee air traffic controller. In a training centre such as the Transport Canada Institute in Cornwall, there are a number of pseudo-pilots who assist in the training of controllers. There are many reasons for developing an automated pseudo-pilot:

- the costs (training and wages) of using human pseudo-pilots;
- the shortage of qualified human pseudo-pilots in general and when a trainee controller feels that he or she needs to practice a specific procedure;
- the possibility of improving the standardization of training given to controllers.

Under the terms of the agreement with Transport Canada, this study examines the following aspects in turn:

- The current state of applied research in the field of automated pseudo-pilots. For convenience's sake, we look only at applied research in North America and Europe.
- The development of a "custom-made" model of the operating language used by trainee controllers, with a view to maximizing the performance of a speech recognition system. This step involves selecting the air traffic control situation to be used as the basis for modelling language. Note that the purpose here is not to develop a speech recognition system or strategy; rather, we are trying to create a representation of the operating language used in real life for application in certain exercises that can be handled by a speech recognition system.

- Speech recognition and generation technologies. We look at the requirements of the task, ergonomic considerations, and the characteristics of the technologies available.

Although this study suggests answers, we hope that it will also raise fundamental questions concerning:

- The role that an automated pseudo-pilot is intended to play in controller training.

- The advisability of developing a 100 % Canadian pseudo-pilot.

2. THE CURRENT STATE OF APPLIED RESEARCH

This chapter sums up the current state of applied research in North America and Europe, as concerns the design of automated pseudo-pilots (including certain related commercial products). We examine a number of automated pseudo-pilots that have been developed by private companies, governments, or a combination of the two. In particular, we look at the modelling of the language of trainee controllers, the ergonomics of the dialogue with the pseudo-pilot and existing prototypes. We make no attempt here to evaluate the speech recognition systems used, for the evaluation of speech recognizers is still at a primitive stage. In addition, the speech recognizer is only one component of an automated pseudo-pilot. Accordingly, one must evaluate the performance of the system as a whole (including, for example, the language model), rather than the performance of a single component.

2.1 Canada

We identified only one organization that is developing simulation systems using voice input for controller trainees, namely ATS Inc.[1] This company has already played a role in the development of a simulation system for trainee controllers (1).[2] More recently, ATS developed a system that can handle voice input. According to our information (2), the system is used to train controllers in radar control situations. The system runs on a PC 80286. For speech recognition, it offers the following features:

- Verbex connected-word speech recognizer;
- the recognizer uses finite-state syntax, i.e. a tree

---

[1] ATS Aerospace Inc., 3163 Harvey Street, St. Hubert, Quebec, J3Y 3T7, (514) 676-6259.

[2] Numbers in parentheses refer to the list of references at the end of the report.

structure describing the possible "routes" for a syntactically valid sentence. The vocabulary contains about 300 words.

For speech generation, the system can use either a synthetic voice or a digital one. According to the same source, ATS has not yet installed this kind of system (with voice input). There is, however, a demonstration system with this option.

2.2 United States

Various civil and military government organizations are interested in developing an automated pseudo-pilot or related systems. In this study, we concentrate mostly on research in the civil aviation field, except in cases where it is worth examining systems designed on the basis of military research.

The FAA (Federal Aviation Administration) is the civil organization that oversees air transport in the United States. The FAA also numbers various organizations interested in controller training and the use of voice technologies for developing automated pseudo-pilots.

Within the FAA, there is general agreement (3) on the need to improve the amount of simulation time (and hence the number of pseudo-pilots) available for trainee controllers to meet anticipated needs. Nonetheless, there are budget and staff constraints, making the alternative, i.e. automated pseudo-pilots, attractive. In the FAA, it is felt that an automated pseudo-pilot can be used at various steps in training a controller (4): at the initial stage (the FAA Academy in Oklahoma City), at a more advanced stage (in regional training centres) and even during continuing training of experienced controllers. There is, however, one restriction relating to training in the regional centres: a FAA regulation now in effect states that the actual operating software must be used for the training of controllers. A stand-alone automated pseudo-pilot station thus could not be used, with the existing regulation. Despite this restriction at regional centres, it is planned to evaluate different kinds of automated pseudo-pilots. The FAA is now pursuing two different courses for the development of voice-activated pseudo-pilots:

- a system initially developed for the U.S. Navy by

Logicon;[1]

- an expert system developed by UFA,[2] used to train student controllers, and complemented by voice technology.

These two systems will be described in greater detail below.

2.2.1 Systems developed by Logicon

Logicon has been involved for many years with the U.S. Navy in developing various pseudo-pilot systems employing speech recognition (5), (6). In 1985, the U.S. Navy called for new tenders with the following objectives (ASATS/SATS project):

- develop an automated pseudo-pilot;
- improve standardization of air traffic control training.

The contract was obtained by Logicon. The system is an automated pseudo-pilot used in radar control. Shortly thereafter (1987), the U.S. Navy issued a second call for tenders to launch the TOTS (Tower Operator Training System) program, and Logicon again obtained the contract. The FAA has purchased a TOTS system for evaluation, which Logicon intends to deliver in about December 1990, and the system should be operational in 1991. The system runs on SUN stations, and is described in more detail in (7).

Although these are two different systems (tower control for the TOTS system vs radar control for the ASATS/SATS system), they have some common characteristics:

- The speech recognition system used is the VRS 1280, from ITT. This is a speaker dependent continuous speech recognition system, with a potential vocabulary of 500 words (words can be repeated in syntax for a total of 2000 words). The same card can be used to generate an

---

[1] Logicon Inc., 4010 Sorento Valley Blvd., San Diego, CA, 92121, (619) 455-1230.

[2] UFA Inc., 335 Boylston St., Newton, MA, 02159, (617) 964-5172.

artificial voice (pre-recorded with a CVSD compression algorithm).

- The speech recognition system uses a finite-state syntax.

According to the supplier of the speech recognizer used, these training systems also have a number of difficulties in common (8). A number of them are relevant to civil aviation applications:

- There may be several instructions included in a single transmission for a given aircraft.
- There may be different messages included in a single transmission, intended for different aircraft.
- Some transmissions are fairly lengthy (30 to 40 words).
- There must be an attempt to handle the corrections that trainee controllers issue in the messages.

The requirements of the U.S. Navy for speech recognition (9) are 98% for words, and 94% for sentences. According to internal evaluations at Logicon, the system now has a 96.4% recognition success rate for sentences. This performance level is in part attributable to internal processing in the pseudo-pilot to correct certain recognition errors.

2.2.2 System developed by UFA

The UFA company has developed a system (ATCoach) that can be used for training air traffic controllers (10). Work on the system began in about 1984, following a call for tenders by the FAA. This is basically a training system for air traffic controllers that relies on simulation and artificial intelligence techniques. The trainee controller may communicate with the system either manually or orally. The system is illustrated in the figure below (adapted from (11)).

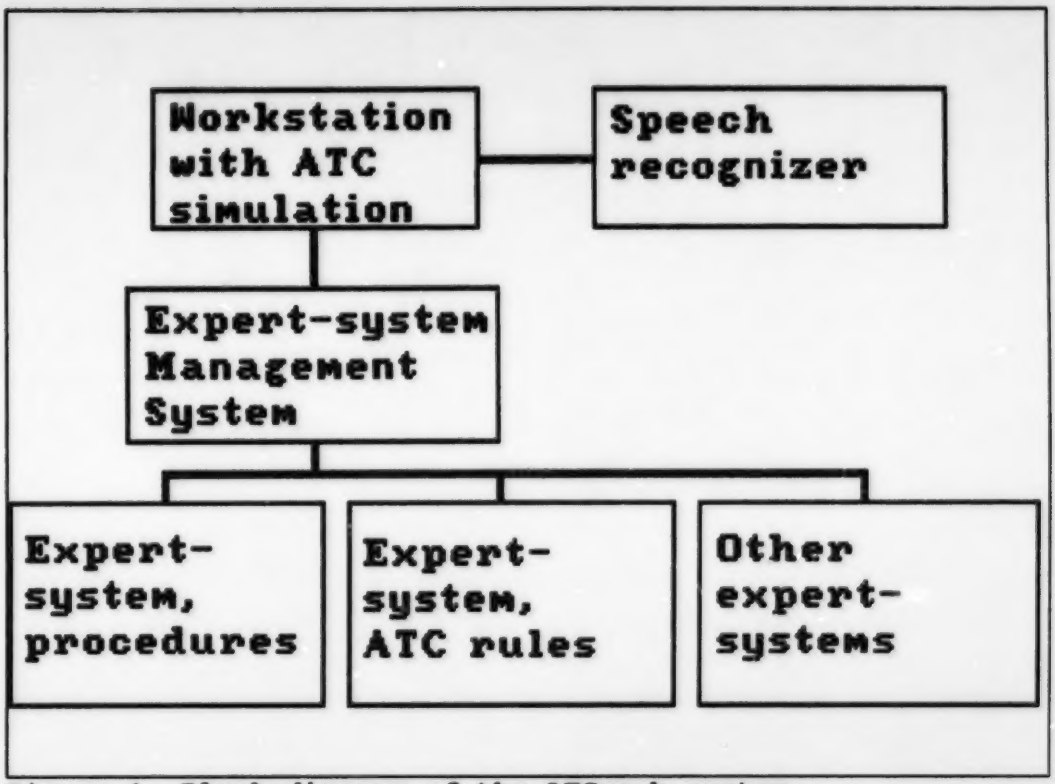


Figure 1 **Block-diagram of the ATCoach system**

During an exercise, the system receives instructions from the trainee controller and evaluates them before carrying them out. One of the expert systems notifies the trainee if the instruction violates established procedures.

It is possible to have a stand-alone station, or a network of several stations to allow hand-offs among control sectors. The platform used for the simplest version of ATCoach is a PC with an 80386 processor and 33 MHz clock speed. More elaborate versions can run on Apollo workstations (12). A pre-recorded digitized voice is used for the automated pseudo-pilot.

At the request of the FAA, two separate speech recognition systems have been tested for the ATCoach system.

. The speech recognition system produced by Verbex, which operates in speaker-dependent mode using finite-state

syntax. This system accepts connected words, and has a vocabulary of about 300 words. Speech synthesis uses a digitized voice. We were unable to obtain performance data of this system with the automated pseudo-pilot.

. There is also a speech recognition system produced by Speech Systems Inc. (SSI) of Tarzana, Ca. This system also uses a finite-state grammar and is reportedly a speaker-independent system. Nevertheless, it is possible to adapt (or "profile," as the company terms it) the speaker-independent system for users who obtain poor results with it. The system is based on neural networks and operates with connected words. The system used thus far was trained with words taken from everyday vocabulary, rather than the vocabulary specific to air traffic control, but that training is planned as part of a development project (now in progress) carried out for the FAA. The following table illustrates some of the results available (13):

Table I **Performance of the SSI speech recognizer for the ATCoach system**

Mode	error rate (%)	
	words	sentence
Speaker-independent	6.7 %	28.5 %
Speaker-independent, profiled for the speaker	1 %	10 %

Although the performance of this speaker-independent system can be considerably improved by profiling it for a speaker, there is

still an appreciable error rate for sentences (10%).

2.2.3 ATC phraseology training system

ITT has produced an ATC phraseology training system, with the following features (14):

- Speaker-dependent connected word recognition, using the ITT VRS 1280 speech recognizer. The platform is a PC-AT 286 running under DOS. The programming language is C.

- The system uses a "flexible" finite-state syntax, making it possible to recognize sentences with certain legal syntactic constructions, as well as illegal constructions. Semantic blocks (call signs, authorizations, etc.) that may be encountered are included at the same level in the syntax. Thus a sentence may be recognized although its elements are inverted.

The training consists in comparing recognized sentences with scenarios involving the desired semantic blocks. A set of rules, written in C, is used to compare what was said with what should have been said. If errors are detected, the student is advised.

2.3 Europe

This section is based on information compiled from the literature, through tours in France and England, and during a conference organized by the Eurocontrol Institute in Luxembourg in October 1989. The two main players in Europe in developing an automated pseudo-pilot appear to be France and England. A description of the status of applied research in those two countries follows.

2.3.1 France

French researchers have been studying different tasks related to air traffic control for a number of years. Various studies

done on the language used by air traffic controllers in real-life situations have made it possible to identify important characteristics of the language used by controllers (15). Based on some of these studies, the Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI), attached to the Centre National de Recherche Scientifique, in co-operation with the Centre d'Études de la Navigation Aérienne (CENA) and a private-sector partner have prepared a design for an automated pseudo-pilot (16). The pseudo-pilot is considered a complete system, designed to replace a human. The trainee controller interacts completely with the system, in the same way as he or she would interact with a human pseudo-pilot. The objective of the project, then, is to produce a robust and user-friendly system. It is agreed that the system must:

- recognize different ways of phrasing the same message;
- handle errors by the speech recognizer or the trainee controller.

The approach employed at LIMSI integrates knowledge related to the task, vocabulary, phraseology and the history of the dialogue in a hierarchy of frames (17). The figure below briefly describes the system (adapted from (18)):

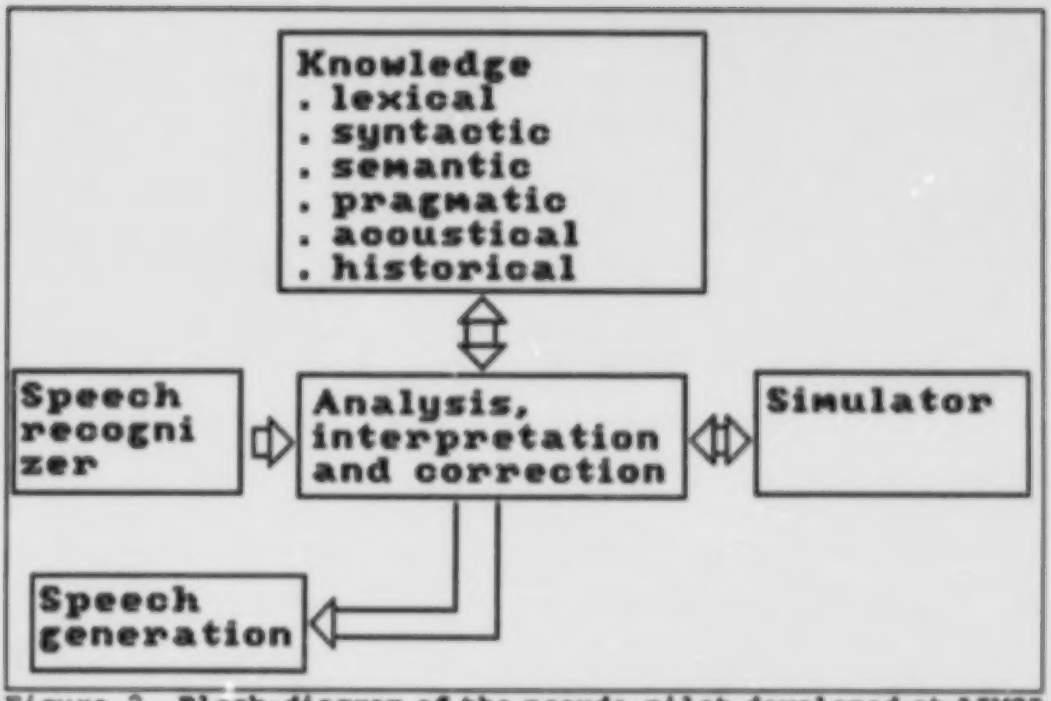


Figure 2 **Block-diagram of the pseudo-pilot developed at LIMSI**

. The trainee controller issues a message, which is recognized by a speech recognizer. In an earlier version of the system, the speech recognizer operated with free syntax, that is, it could recognize all the words in the vocabulary at any time. As the system evolved toward more complex tasks, a sort of flexible syntax became necessary, which both improved the system's performance and allowed the trainee controller great freedom of expression. The flexible syntax in use, termed "evolutive binary syntax," changes with the dialogue, and involves defining for each word in the vocabulary a list of words that may follow it. A more complete discussion of the evolution of the syntax is found in (19).

. That message is associated with a frame, using a word pairing strategy for the words in the message and the words in a frame. The elements of the message are then associated with the slots in the corresponding frame (instantiation).

- The system then proceeds to detect and correct any errors. A message must be consistent with earlier steps in the dialogue and with itself (for exzmple, an aircraft cannot climb to an altitude lower than its current altitude). Various strategies are used to correct errors without having to call for intervention by the trainee controller. Only if the automatic correction fails is the user asked to repeat part or all of the message.

- The dialogue is then updated to take account of the contents of the message, and the appropriate action is taken in response to the message (e.g., echo the message from the trainee controller and change headings).

- Finally, the system attempts to predict the type of message that the trainee controller will issue next, with a task model and a user model. The predictions are used to alter the syntax and simplify the task of the speech recognizer.

The system is currently a demonstrator, with the following features:

- The speech recognizer is speaker-dependent, and recognizes connected words. The Datavox system developed by LIMSI and marketed by Vecsys is being used at the moment.

- The language model used is frame-based, with the various types of messages likely to appear in the dialogue associated with the frames.

- The voice of the pseudo-pilot uses word concatenation. The system incorporates a model of the environment for each aircraft and builds a dialogue history.

- The platform is a PC 386, and programming is done in LISP and C.

According to the results available (17), with a vocabulary of 110 words and an evolutive binary syntax, the system achieves a recognition rate of 99.5%. This rate is based on correct semantic comprehension of the message, regardless of errors by the speech recognizer.

2.3.2 Great Britain

There has been interest in applying speech recognition technology to the field of air traffic control for many years in this country (20), (21). Although no applications are foreseen in the short term for voice technology in the operational field, it is planned to evaluate and use that technology in roles for which it is considered suitable at present. Accordingly, a study of human aspects of speech recognition and the phraseology used by air traffic controllers in actual communications has been sponsored (22).

The British seem to have taken a different approach from the French with regard to the design of an automated pseudo-pilot, at least in the short term. Rather than an independent system for training novice controllers, the British have developed a hybrid system using both voice and touch input. This system is used for exercises in which experienced controllers participate. Those controllers do not need to train the system in air-traffic control phraseology or procedures, and use speech as the primary input method for giving instructions to the automated pseudo-pilot, turning to a touch screen if a command presents difficulties. The system replies with a synthetic voice (Dectalk).

An improved version of this pseudo-pilot is planned for 1990. It will use a speech recognizer (possibly Macrospeak), but the possibility of manual input via the touch screen will remain. An interesting feature of this improved station will be that the acquisition of voice samples necessary for training the speech recognizer will be transparent to users. Existing portable units used to provide complementary training for controllers will be employed to gather voice samples. Samples from a given controller will be loaded into the speech recognizer of the pseudo-pilot when the controller in question begins a simulation session.

2.4 Summary of current applied research

This review of research is not exhaustive. Much of the research in this field is very practical, often carried out by private-sector companies that do not publish their findings. Thus it is possible that certain systems have not been mentioned here. For the systems examined thus far, we can nevertheless

make a number of observations. First of all, note that the different systems have substantial differences in terms of:

- the air traffic control situation for which an automated pseudo-pilot is designed (tower control vs radar control);
- the language model used (finite-state syntax vs. frame-based model);
- the role of the automated pseudo-pilot (simple simulator vs. advisor in carrying out tasks).

Note also that there has been no complete evaluation of these systems. We have a certain number of results concerning prototypes, but since none of the voice-driven pseudo-pilot systems is yet in use in a training centre, there has been no overall evaluation of their performance, and it is impossible to determine whether all the objectives have been reached. In addition, discussions with the designers of some of these systems show that many choices were made intuitively, with no attention to ergonomic considerations. The choice of a pre-recorded digitized voice for the automated pseudo-pilot, for example, is essentially based on a subjective evaluation of the quality of the speech generated, rather than the result of a rigorous evaluation.

According to our observations and in line with the objectives described earlier, we are now ready to move on to the second part of this project: the design of a custom-made model of the language used by trainee controllers. That involves the following steps:

- deciding for which air traffic control situation we wish to model the language;
- selecting the formalism to be used to design the model;
- determining the approach to be taken in designing the model;
- designing the model.

The next chapter begins the process, by selecting the air traffic control situation for which we will develop the language model.

3. SELECTING AN AIR TRAFFIC CONTROL SITUATION FOR THE DESIGN OF A LANGUAGE MODEL

The purpose of this chapter is to select an air traffic control situation suited to the design of a model of the language used by trainee controllers. First of all we deal in general with the phraseology used in different air traffic control situations. We then select one of those situations as the subject of this study, and describe it in greater detail. Finally, we give an overview of the physical environment of the position studied. This step precedes the analysis of the language used by trainees which is covered in a later chapter.

3.1 Phraseology used in different air traffic control situations

We have taken a simple approach, first dividing air traffic control into "visual flight" and "instrument flight" situations. We then examine the phraseology used in each and select the situation best suited to the development of a language model.

3.1.1 Visual flight control

Air traffic is controlled under visual flight conditions with no radar coverage and with the controller in direct visual contact with the aircraft. The instructions issued by the controller to the pilot refer to landmarks easily visible from the aircraft. Thus the phraseology is much less standardized under visual flight rules than under instrument flight conditions. Discussions with the course directors at the Transport Canada Training Institute confirm that in some cases, the official phraseology is insufficient, and must be complemented with natural language. Modelling the language in a form that can be used by a speech recognizer then becomes quite difficult. In addition, an automated pseudo-pilot for VFR air traffic control must incorporate a realistic visual representation of what a controller might see from a control

tower, which considerably increases the complexity of designing an automated pseudo-pilot. For those reasons, we decided not to use a VFR situation for developing the language model.

3.1.2 Instrument flight control

According to course directors at the Transport Canada Training Institute, the vocabulary used in IFR air traffic control is highly standardized. In addition, students at the Institute are encouraged to stick fairly closely to that phraseology. Thus we will analyse this type of situation to see how an automated pseudo-pilot can be designed. We must also distinguish between IFR air traffic control with and without radar coverage. An examination of the documentation describing the official phraseology gave no indications of the advantages of either type, and so we arbitrarily chose instrument flight control with radar coverage.

3.2 Description of the situation considered in this study

The type of control conducted under IFR conditions with radar coverage depends on altitude and the sector of the airspace considered. The following figure shows a cutaway of a fictitious airspace used for exercises at the Transport Canada Training Institute in Cornwall, part of the regional control zone for Midtown (24).

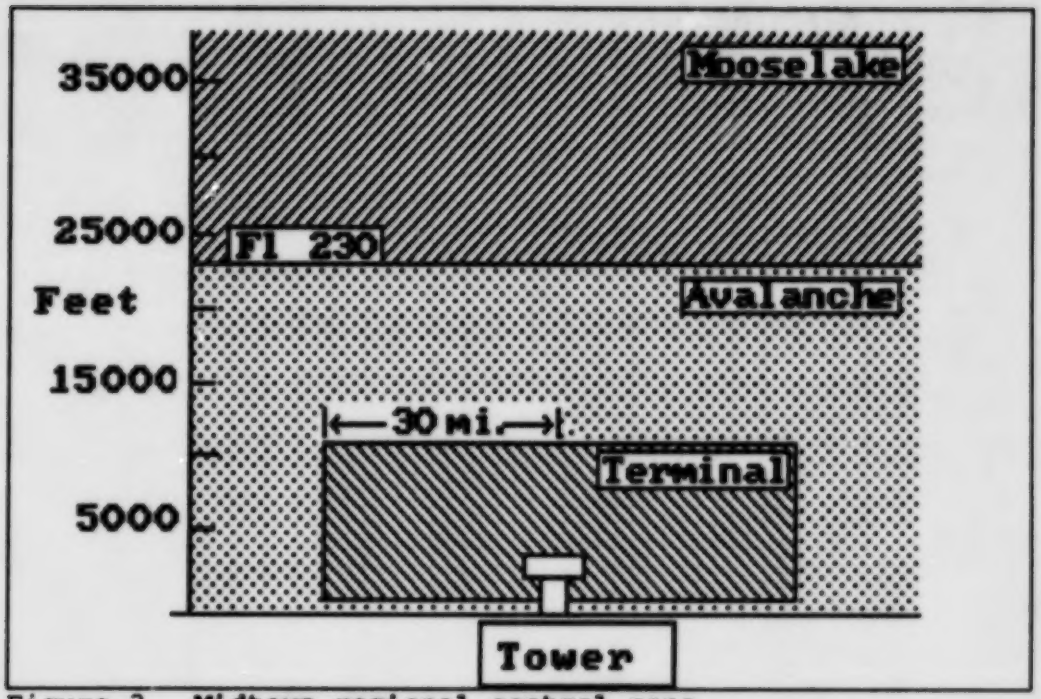


Figure 3 **Midtown regional control zone**

The scale on the left indicates altitude. By convention, altitudes up to 18,000 feet are given in thousands of feet (e.g., "eleven thousand feet"). Altitudes at or above that level are given in "flight levels": a flight level refers to a multiplier applied to a height of 100 feet (e.g., "flight level one eight zero" corresponds to 180 x 100 feet, or 18,000 feet).

It can be seen that the Midtown regional control zone is itself divided into several zones. The following zones are used for the exercises recorded at the Transport Canada Training Institute in Cornwall:

- A high-altitude control zone: the Mooselake zone takes in all altitudes above flight level 230.
- A low-altitude control zone: the Avalanche zone comprises all altitudes from the ground up to flight level 230.
- A terminal control zone: the Stilpond terminal provides the interface between the airport and other zones, for

takeoffs and landings.

The characteristics (radio frequencies to be used, routes, etc.) and geographic boundaries of these zones are defined in the reference documents issued to trainee controllers. In addition, the procedures specific to each regional zone set out the protocol to be observed for handing off an aircraft within a regional zone. For example:

- There is a preferred heading and altitude for handing off an aircraft between the high-level zone (Mooselake) and the low-level zone (Avalanche).

- The process of handing off an aircraft to another zone must be initiated at specific times while an aircraft is en route toward another zone.

There are similar procedures among the various regional zones. In this case, they are stipulated in the agreements between regional zones.

To select the part of the instrument flight control task to be used, a number of factors must be considered:

- The complexity of the task: the best approach is to first try modelling a simple task.

- The sequence followed by trainee controllers during their simulation training: they are given step-by-step training that familiarizes them with all the control procedures used in the different zones, one at a time.

According to the course directors at the Training Institute, it appears that controlling air traffic in the terminal zone is the most comprehensive task (i.e. the one that involves the most procedures). Since it also a fairly complex task, we adopted the following approach:

- We examine a simple task, from which we draw a basic, but complete model. To do this, we make recordings of an elementary exercise held in a low-altitude control zone. During this exercise, the controllers are at the beginning of their training in instrument flight control.

- We then attempt to adapt this model to a more complex task, in this case terminal control. Controllers at this

point are nearing the end of their instrument flight control training.

This progressive approach thus corresponds to the course plan used at the Training Institute (see the description of the contents of Segment Two, in Appendix A).

3.3 The trainee controller's workstation

3.3.1 Physical features

The following figure shows a sketch (not to scale) of the trainee controller's workstation.

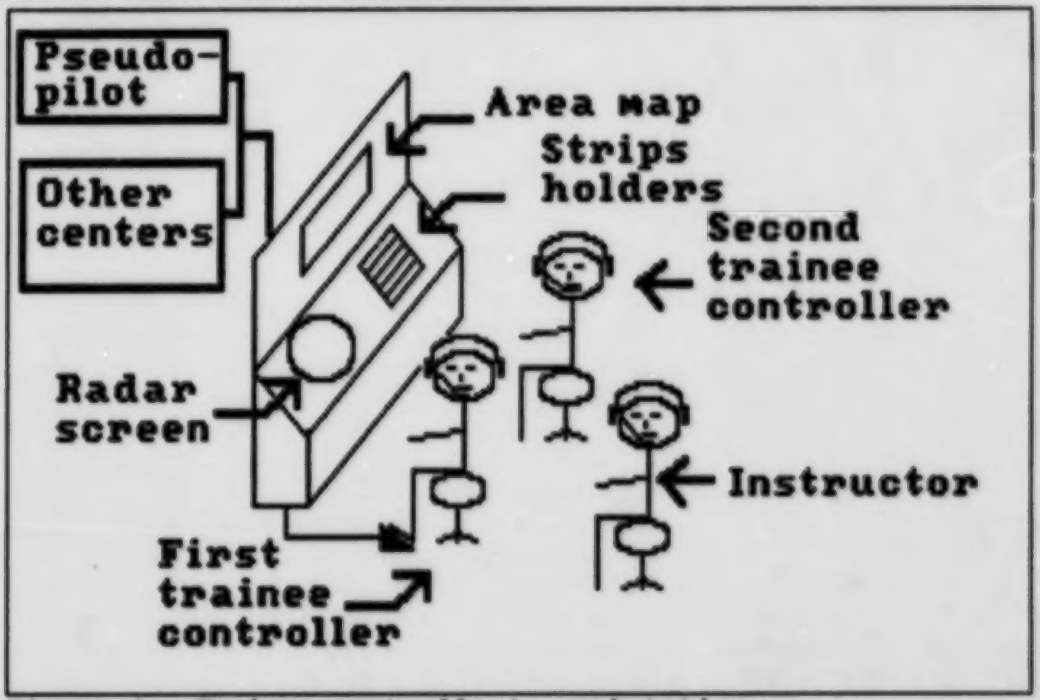


Figure 4 Trainee controller's workstation

The trainee controller and the pseudo-pilot are physically located in different parts of the simulation room.

3.3.2 General description of communications

Only some of the trainee controller's communications are relevant for this study. They may be classified as follows:

- Communications with the instructor, who is generally located right behind the trainee. The trainee controller may request or receive advice or information from the instructor. In addition, the trainee may also communicate with a second trainee controller at the same workstation. The second trainee ensures co-ordination with other centres, among other tasks, and does not intervene in communications with the pseudo-pilot. Communications among the two trainee controllers and the instructor involve no communications devices.

- Communications with other centres. These communications are conducted using the same headset as is employed for communications with pilots. The same channel is also used, with one difference--a manual switch is used to transfer the call to other centres. In real air traffic control situations, the pilot does not hear these communications.

- Communications with the pseudo-pilot. These communications are conducted through a headset. A dedicated foot-operated switch allows the trainee to speak to the pseudo-pilot.

3.3.3 Observations on training procedures

One exercise lasts from 35 to 40 minutes. Before beginning the speech exercise as such, the instructor explains the various possible situations, using overheads. Trainees then take their places, and the vocal exercise begins.

There are a number of circumstances that may influence the performance of a speech recognition system:

- During the exercise, the trainee controller in charge of the workstation answers requests from pilots, or initiates conversations. Different trainees wear their headsets in different ways: some wear them correctly

and apparently do not tend to vary the position, while others move their headsets around during the exercise.

- There are sometimes conversations close to the trainees while they are conversing with the pseudo-pilot.

- Controllers are obliged to manage information by taking notes on paper ("flight strips"). These strips are inserted in holders and the holders are moved around, handled and so on. The handling of these holders produces frequent short and sharp noises. The integration of speech technology should take this into account.

- During our observations, it became apparent that a controller was speaking with a single pseudo-pilot who played the role of a number of pilots. Consequently, the controller always hears the same voice. The pseudo-pilot may also play the role of a controller in another centre who is in contact with the trainee controller.

One more step remains before beginning the language modelling process, and that is to select the formalism to be used to represent (or model) the language. That step is dealt with in the following chapter.

4. SELECTING THE MODEL TO REPRESENT CONTROLLER-PILOT LANGUAGE

This chapter is aimed at choosing a model to represent the language used during trainee controller and pseudo-pilot conversations. The models used thus far in the design of automated pseudo-pilots are analysed, to arrive at a choice. By a model of the language we mean a formalism describing:

- the actual lexicon of the operating language;
- the structure of the operative language.

This model is necessary to allow a speech recognition system to recognize sentences, by providing a lexicon and, if necessary, the syntax of the language. In addition, our understanding of the task will improve as the model is developed. Finally, the model will also be useful for defining requirements with regard to the speech recognition and generation system.

4.1 Analysis of existing models

The work carried out thus far on the application of speech recognition in air traffic control has taken two approaches to modelling the operating language used in air traffic control (8) (13) (14) (15) (18) (25). Each of those approaches is based on a different school of thought. There may be:

- a "strict" philosophy, calling for precise use of the syntax and restricting the number of formulations allowed for communicating a given concept;
- a "flexible" philosophy, allowing various formulations to express a given concept and tolerating some divergence from a given syntax (e.g., abbreviations, inversions, etc.).

We will examine the models corresponding to each of those philosophies.

4.1.1 Representation using a syntactic tree

This representation comprises the lexicon and syntax of the language in a tree structure. A syntactically valid sentence made up of words belonging to the lexicon is represented as a series of branches in the tree. The following figure illustrates the use of a (hypothetical) model of this type to recognize the sentence "turn left heading 355."

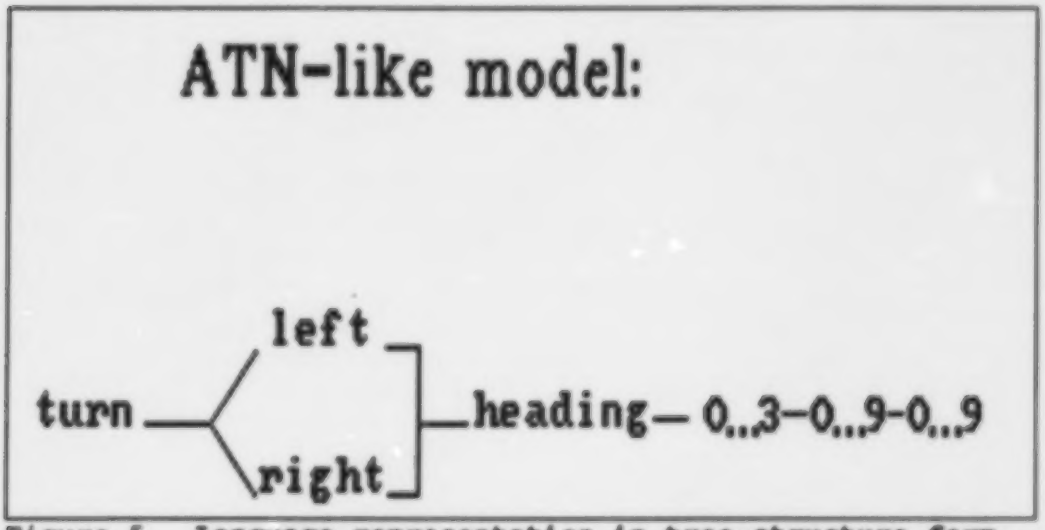


Figure 5 **Language representation in tree-structure form**

This approach makes it possible to reduce the number of words likely to follow the last word recognized, and thereby considerably improve the performance of speech recognizers. This representation has been the most used since the introduction of speech recognition. At first glance, since this model covers the entire lexicon and the allowed syntax, it ought be possible to describe all the legal spoken exchanges likely to occur. That might be the case if sentences were always perfectly assembled during an air traffic control task. But the analysis of many examples taken from conversations between pilots and controllers in real-life situations (15) (22) shows that:

- There are various formulations (either officially sanctioned or founded on usage) for expressing the same concept.

- There are also a number of syntactic "errors": pauses, hesitations ("uh," "er," etc.), additions ("good day"), repetitions, and so on.

If we wish to use a tree structure to model the language, while maintaining the flexibility and transparency of the dialogue, we must include several different formulations for the same concept (i.e., add branches to the tree, or extend existing branches). That has two undesirable consequences, however:

- It increases the branching factor in the structure: the branching factor corresponds to the number of possible branches following a given word, and a high factor makes it more difficult for the speech recognizer to select the proper branch.

- It may quickly become awkward when the time comes to train the speech recognizer.

Even if these two difficulties can be overcome, other problems are not far behind. For example, the use of an unknown word may make it impossible for the system to recognize an otherwise completely legal sentence.

It is possible to try to overcome the problem of errors (pauses, hesitations, repetitions, etc.) by trying to predict where certain errors are likely to occur and to add corresponding elements (i.e. pauses, "uh"s, etc.) in the model to compensate. That leads to another difficulty:

- It is difficult to predict exactly where dialogue errors will occur. The rigidity of the tree structure representation makes it essential to model exactly where each error will occur.

Adding words in the syntax (e.g., "correction") to compensate for errors also leads to a heavier syntactic tree. In addition, with this model there is no simple way to introduce a means of taking all possible repetitions and corrections by the speaker into account.

4.1.2 Frame representation

This approach relies on the fact that verbal exchanges in air traffic control can be associated with an operative language and classified in a limited number of categories. Those categories are then associated with frames. A frame is a data structure consisting of a series of slots, which describes a stereotyped situation (26) (in this case, types of verbal exchanges). These slots may contain references to other frames, conditions, default values or procedures. Such a model for representing knowledge generally comprises several frames (at least one frame per category of message in the language). For purposes of air traffic control, this approach has been used in (15). The following figure illustrates the use of a model of this type for recognizing the sentence "turn left heading 355".

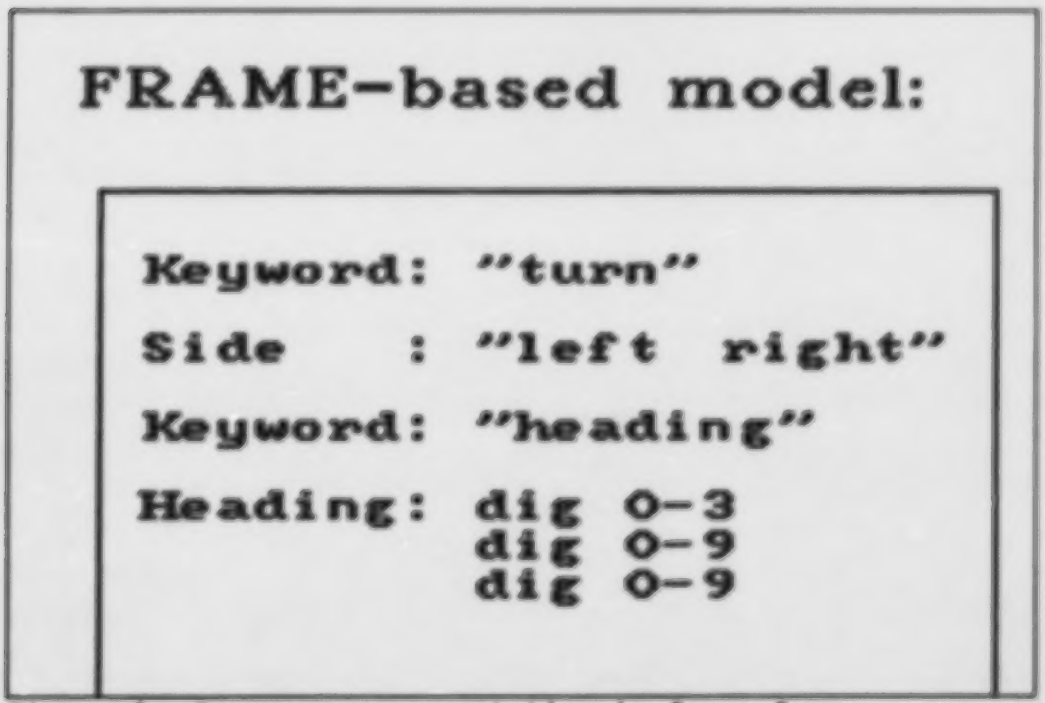


Figure 6 Language representation in frame form

It can be seen that this type of representation is fundamentally different from the syntactic tree described earlier. While the latter analyses a sentence by using

syntactic constraints, the frame-based model tries, instead, to associate a message with a given context (the figure shows the frame relating to an instruction to change headings). Once the context has been identified, the next step is to pair off the elements of the context and the contents of the message (i.e. by identifying the side, "left" or "right", as well as the heading, "heading 355"). This representation concentrates on the semantics of the message; there is no prior decision concerning the order in which the different elements of the sentence are to be found. It also makes it possible to use a form of "local" syntax (i.e. the presence of syntactic nuclei inside the sentence is recognized, without their location being strictly fixed). Using the model to understand the message thus basically involves two steps:

- First, recognize the category to which the message belongs.

- Second, instantiate the contents of the message in relation to the expected contents.

The frame representation makes it possible to elegantly resolve the difficulties caused by the existence of a number of formulations for a given concept. Indeed, the instantiation of the frame allows the key elements of the message to occur in a different order. In addition, "errors" are simply accepted as such: events likely to occur in certain situations, but difficult to predict accurately. In that case the fact that certain frame slots are optional is an advantage.

A frame-based language model integrates naturally into an automated pseudo-pilot system using a declarative approach. It is possible to use a uniform object-oriented formalism to represent knowledge of the task universe (control zone, aircraft, etc.) and make the automated pseudo-pilot more intelligent.

There are also some inconveniences involved in this representation, including:

- To be able to use this formalism, the language to be modelled must be suitable. Message categories with different meanings must be sufficiently distinct from one another for a message to be classified properly with no ambiguity.

- The learning process by certain speech recognition systems attempts to take account of co-articulation between the different elements in the lexicon. If all the words are allowed to follow other words in any order (i.e. without syntax), training the speech recognizer becomes problematic. It may be advantageous to use a "flexible" type of syntax that makes it possible to improve the recognition of certain elements, while simplifying the training process for the speech recognition system.

It is also possible to try to generate an evolutive syntax, depending on the context and status of the dialogue at a given time. This method is in fact used in an existing prototype (19).

In another connection, the choice of frames can be restricted to a limited number of frames that are plausible in the given situation. This amounts to creating a relatively rigid "frame syntax," or a script. A script is a structure that describes an appropriate sequence of events in a particular context (27).

4.2 Selecting the model to represent the language

The selection of the model depends on a number of factors, in particular whether the "strict" or "flexible" philosophy is to be applied. The choice also depends on the level of training that the automated pseudo-pilot is to be used for. Totally new controllers can probably benefit from very strict phraseology in their training, but for more advanced students it is likely that a richer vocabulary would be more suitable (as is now the case with human pseudo-pilots).

Where this study is concerned, the frame-based representation model will be used. We feel that we will be able to obtain a more flexible model, imposing additional syntactic constraints at a later date if a very strict phraseology is considered necessary. In addition, we feel that the use of techniques drawn from artificial intelligence will be essential sooner or later in an automated pseudo-pilot application. Using frames to model the language will allow us to apply these techniques more easily.

We are now ready to explain the approach to be used to model

the language used by trainee controllers for the air traffic control situation selected.

5. MODELLING THE OPERATIVE LANGUAGE

This section explains the approach used to create a custom-made model of the operational dialogue employed by trainee controllers. It was agreed to conduct the study for dialogues in English. The model must make it possible to describe exchanges between trainee controllers and pseudo-pilots in a way that can be used by an automated pseudo-pilot. We have the choice between modelling either the official phraseology recommended in (23), or the phraseology actually used by trainee controllers. We decided to model the language actually used by trainees, for the following reasons:

- As noted in Chapter 4, there are in fact differences between the official phraseology and the phraseology actually used by controllers. An automated pseudo-pilot using a model of the actual language will be more "transparent" for the trainee controller, and the exercise will better reflect actual conditions.

- A model is a simplification of reality. Even if we tried to model the phraseology actually in use, it is doubtful that we could do it justice. Accordingly, even a model of the phraseology actually used is more restrictive than the real phraseology. Thus we arrive at a happy medium between the official phraseology and that in actual use.

Since the language model must be used to understand the meaning of the instructions given by trainee controllers, the analysis of the language will consider mostly the messages issued by those trainees. We feel that the answers by the pseudo-pilot will depend directly on the instructions received, in most cases; accordingly, the analysis of those replies is of only secondary importance here.

5.1 Target dialogue

As mentioned in Chapter 3, the expected complexity of developing the model implies that we should begin by modelling a simple situation. We will first model control exercises at

low altitude. We will then consider how the model obtained can be modified to adapt it to a more complex situation, i.e. terminal control.

5.2 The approach to developing the model

In the previous chapter we selected a frame-based model for representing the language. To develop this type of model, we will need the following information:

- the vocabulary of messages: the vocabulary (or lexicon) that the automated pseudo-pilot must recognize in order to recognize and interpret messages;
- the categories of messages: each of the message categories will correspond to a frame;
- the structure of messages: each frame includes a certain number of slots that correspond to semantic nuclei (meaning-bearing elements). These slots (or semantic nuclei) can also correspond to syntactic nuclei (e.g., the means of expressing an altitude). In some cases, minimum or maximum limits, calculation methods, etc., may be associated with these slots.

We decided to take a two-step approach to obtaining those elements. First of all, we conduct a summary analysis of the dialogue, on which a model of the language is based. That preliminary model is then completed during a second step. The following sections describe those two steps.

5.2.1 The first step in developing the language model

The purpose of this first step is to:

- obtain the vocabulary used;
- obtain all the preliminary frames that represent the categories of messages exchanged between the trainee controller and the pseudo-pilot;
- define the contents of the preliminary frames, i.e.

identify the slots belonging to each frame. These slots reflect the semantic elements in each message (i.e. the meaning-bearing elements). In addition, if the model is implemented at some future date, the contents of the slots may act as local syntactic nuclei.

The following figure illustrates the first step in developing the language model.

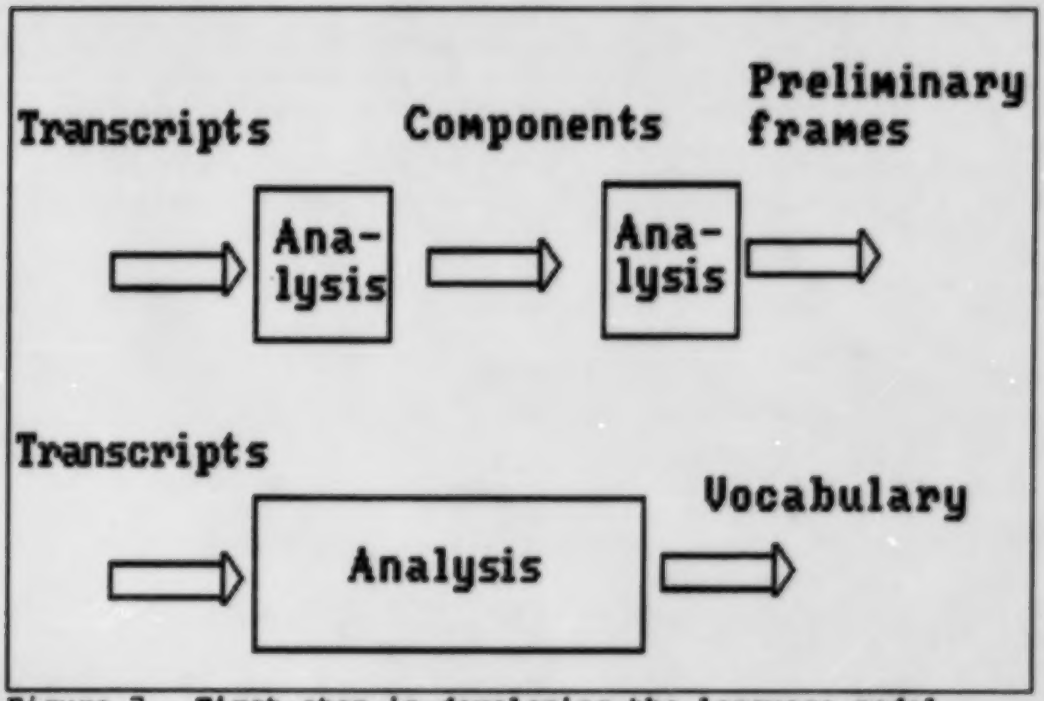


Figure 7 First step in developing the language model

We first examine all the messages from the trainee controller, and identify their components (call signs, altitudes, etc.). Those with characteristics in common (same keywords, same components, similar length) are then grouped together. We do that for each transcript, omitting situations involving dialogue errors which will be examined at a later stage. The results of all the transcripts are grouped together, following any necessary adjustments.

We also establish the vocabulary of messages from the trainee controller for each transcript. The results are then combined.

The development of preliminary frames and the identification of vocabulary can be done independently.

5.2.2 The second step in developing the language model

The preliminary model obtained in the previous step is incomplete, lacking both the mechanisms for instantiating certain frame slots, and error handling mechanisms. In addition, we wish to determine whether it is possible to predict which frames will be called during an exercise. This last step in the process thus completes the preliminary model. The objectives of this step are as follows:

- identify the instantiation mechanisms for various slots in the frames;
- briefly examine dialogue error handling mechanisms;
- identify rules for calling successive frames.

To reach these objectives, we first draw up frame instantiation rules that make efficient use of all available resources. Since the dialogue takes place as part of a well-defined exercise, the knowledge of the "microworlds" of the controller and pilot may be used to overcome the limitations of speech recognition technology. We can supplement the frames with secondary sources of information or default values to be used when the dialogue itself is insufficient.

In conjunction with this process, we determine how errors are managed during the dialogue, and define rules to be used in modelling that part of the dialogue.

Finally, we attempt to determine which frames will be called during a given scenario. We identify rules to help limit the number of frames to be called in analysing a message. We then have an accurate model of the operative dialogue. The following figure illustrates this step.

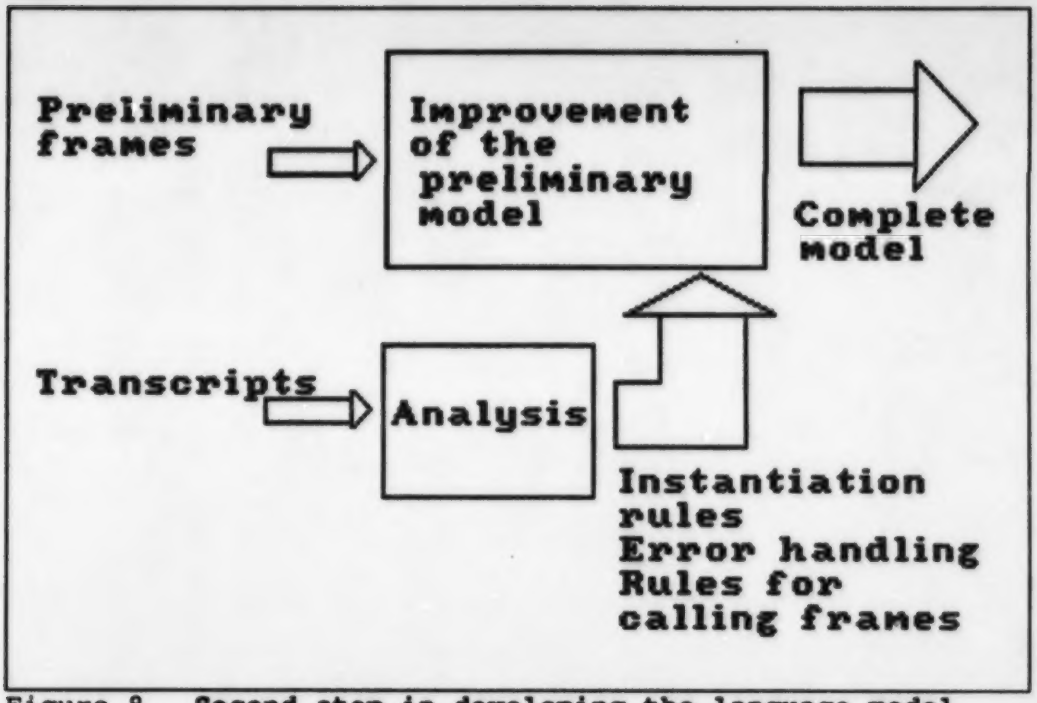


Figure 8 **Second step in developing the language model**

We can now design the model as such. The following two chapters describe the application of this method to modelling the language used in simple low-altitude exercises.

6. FIRST STEP IN DEVELOPING THE MODEL FOR SIMPLE EXERCISES

This chapter applies the method described in the previous chapter to analyse transcripts of real operating dialogue. As mentioned in Chapter 3, we decided to begin by analysing two recordings of a simple low-altitude instrument flight radar control exercise conducted at the Transport Canada Training Institute in Cornwall, Ontario. Each recording lasts approximately 35 minutes, and corresponds to an exercise by a different trainee controller. We transcribed both the recordings, spelling them as accurately as possible. The transcripts are called 1a and 2a, and are reproduced in Appendix B. We had planned to use four recordings, and hence four transcripts, but a technical problem obliged us to eliminate two of the recordings. Despite the smaller sample, we expect to be able to obtain a preliminary model of the operating language; we will later determine whether it is a valid model, given the reduced size of the material on which it is based. To simplify the development of the model, we imposed the following constraints:

- Messages (coherent sets of information belonging to a given category) are not concatenated during a transmission (radio communications).

- A transmission is sent to only one recipient.

These constraints will be lifted later.

6.1 Examining the length of transmissions

Here we will examine the distribution of the length of transmissions issued by trainee controllers. This information will be useful for obtaining an idea of the complexity of processing to be done by the speech recognition system. The following figure shows this distribution, for simple exercises:

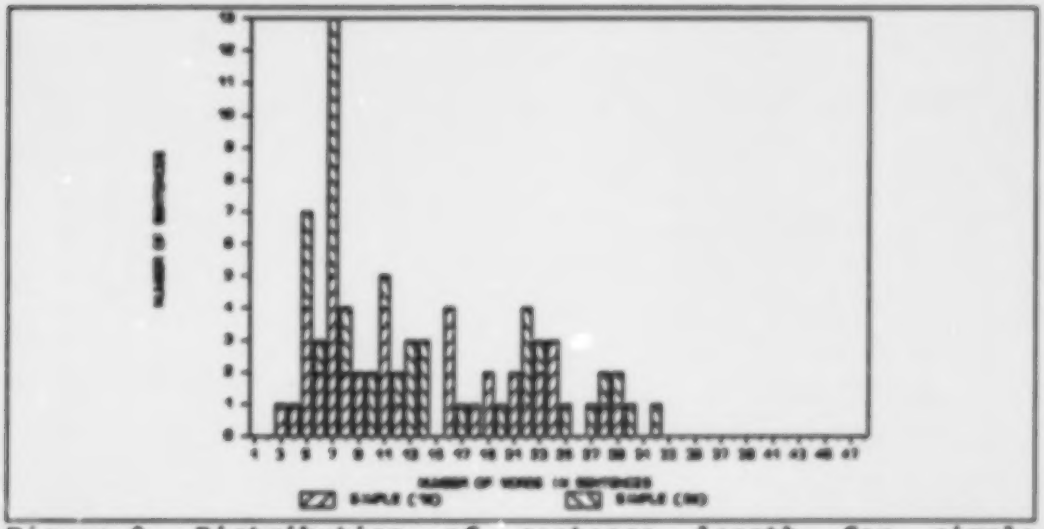


Figure 9 Distribution of sentence length for simple exercises

It can be seen that even for simple exercises, there may be sentences with up to 32 words. Given the size of the sample, we can consider this a lower limit, i.e. the maximum length of a sentence for this type of exercise is at least 32 words.

6.2 Obtaining the vocabulary

In this section we first establish the approximate size of the vocabulary required for an exercise of this type. We also roughly determine the composition of the vocabulary. We know that certain classes of the elements in the vocabulary will depend on the exercise scenario, such as the names of other control centres with which the trainee controller will be dealing, or the names of the different airlines used in the exercises. In addition, the ICAO radio-telephone phonetic alphabet (28) (alpha...zulu) and numbers (in various formulations) may also be used, but their use can vary depending on the controller (e.g., the controller may say "one hundred," rather than "one zero zero"). Since we wish to produce a robust vocabulary, i.e. one that will vary little from one exercise to another, we need a stable list of names of control centres, airlines, radiotelephone phonetic alphabet components and numbers. We will take that into account when obtaining the

vocabulary required.

The following table shows the size of the vocabulary for the transcripts. The complete list of words in the vocabulary of each transcript (simple exercises) is given in Appendix C.

Table II **Results of the evaluation of the vocabulary size**

	Transcript 1a	2a
Total number of words used	74	67
References to control centres	5	4
References to airlines	7	7
References to the ICAO phonetic alphabet	4	4
References to numbers	20	14
Net operative vocabulary (incomplete)	37	37

Each column in the table corresponds to a transcript. For each transcript, the size of the total vocabulary is given. References to control centres, airlines, elements in the radiotelephone phonetic alphabet and numbers have been substracted, giving the net operative vocabulary, to which

certain elements must now be added.

If we are to establish an operating vocabulary, the different users must not only employ a limited number of words, but also use more or less the same words. To see whether that is the case, we combined the net general vocabularies for each of transcripts 1a and 2a to obtain a net combined operating vocabulary of 45 words. This clearly shows that most of the words used were common to both trainee controllers, and we can then suppose that for the same kind of task, the general vocabulary used by several trainees will be about the same size.

As mentioned, to obtain a robust operating vocabulary, we must take account of a set number of control centres, airlines, radiotelephone phonetic alphabet elements and numbers. Using a standard number of those elements makes the vocabulary relatively independent of the specific features of exercises (for example, call signs may vary widely from one exercise to another). Moreover, we can now handle certain individual differences between controllers (e.g., either "one zero zero" or "one hundred" can be used). The following table shows how the required number of those various elements is evaluated.

Table III **Evaluation of the number of extrinsic elements**

Standard element	Method of calculation and source of data	Result
Names of control centres	There are 8 zones that may be mentioned during communications ("Study Notes")	8
Names of airlines	The transcripts show that 10 names seem to be sufficient	10
Elements in phonetic alphabet	(alpha bravo ... zulu)	26
Numbers and variants	(zero one ... niner nine) (ten eleven..... nineteen) (twenty thirty ... ninety) (hundred thousand)	11 10 8 2
Directions	(north south east west)	4
Total		79

The last step in obtaining the size of the operating vocabulary required is to add the combined vocabulary for each type of transcript to the number of standard elements, as shown in the following table.

Table IV **Size of operative vocabulary**

Size of general vocabulary	Number of standard elements	Total size of operative vocabulary
45 words	79 words	124 words

If we add an arbitrarily defined 25% safety margin, the total size of the operating vocabulary is 155 words (124 words x 1.25) for this type of exercise. This margin will absorb the different usages among trainee controllers and will also make it possible to take other elements into account, such as the names of beacons. As mentioned, this is a rough estimate, rather than an exact value.

6.3 Preparing preliminary frames

In this section we show how the preliminary frames for modelling the dialogue are prepared. First of all we examine the transcript of each exercise, creating a temporary frame identified by a name for each message. When we encounter a message that resembles another message (same keywords, same components, similar length) for which a temporary frame already exists, we consider that the new message belongs to the same frame. The following table shows an example of this process, for the simplest frame of those developed. A complete list of individual frames and the messages from which they were derived is given in Appendix D.

Table V Example of the development of a frame

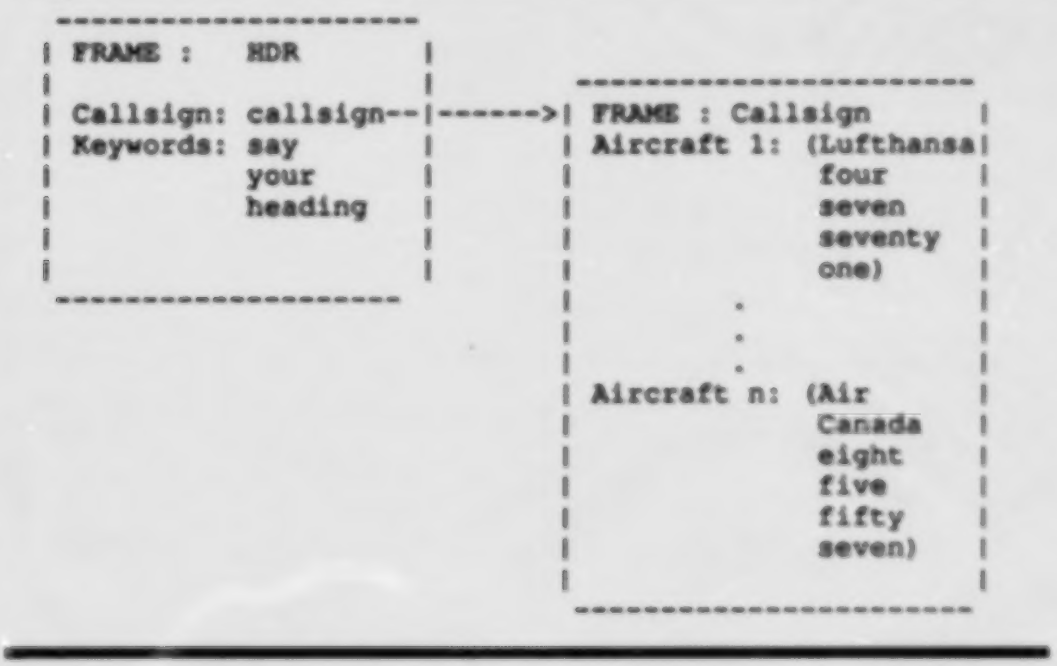
FRAME : HDR

Trans.	Msg no.	Content of message							
1a	1	november	nine	hundred	papa	bravo	say	your	heading
	31	lufthansa	four	seven	one		say	your	heading
	67	alitalia	six	sixty	five		say	your	heading
2a	26	lufthansa	four	seven	one		say	your	heading
	61	alitalia	six	six	five		say	your	heading
		call sign					"say"	"your"	"heading"

For transcripts 1a and 2a, the similar messages (Nos 1, 31, 67, 26 and 61) have been grouped together. The last line of the table shows the fields identified as making up the generic frame "HDR". It can be seen that there is first a variable element (call sign) followed by a certain number of keywords ("say," "your," "heading").

Clearly, it would be awkward to create such a frame for each of the call signs that the controller may refer to. This comment also applies to other parts of the message, such as headings, altitudes, and so on. Thus we find it advantageous to employ a property of the frame method of representing knowledge, that is the possibility that one of the slots of a frame refers to another frame. For the previous example, the only element whose formulation may vary is the call sign. We analyse the way in which call signs are formulated, using the method illustrated in Table 5, and incorporate the results of that analysis in a frame called CALLSIGNS, to obtain the complete preliminary frame shown in the following figure.

Table VI **Representation of frame HDR**



The other frames are prepared in the same way. Thus we obtain a number of frames corresponding to each of the message categories observed in the transcripts. Appendix E summarizes all the frames (and associated sub-frames) for the simple exercises.

6.4 Validating the use of frames in an exercise executed by different trainees

We now wish to verify to what extent the frames prepared describe the messages issued by trainee controllers. In particular, we wish to know whether the model is constant from one trainee to the next. The following table shows the frames used in each of the exercises, as well as the percentage of messages that these frames account for (i.e. the total percentage of messages in the two exercises, excluding messages for error handling).

Table VII **Proportion of messages described by the various frames**

FRAME	Transcript 1a	Transcript 2a	Cumulative, as a % of the total
K	15,6 %	14,3 %	29,9 %
VSD	7,8 %	5,2 %	42,9 %
ALD/F	6,5 %	5,2 %	54,5 %
POD	5,2 %	6,5 %	66,2 %
HDR	6,5 %	2,6 %	75,3 %
VSD	6,5 %	2,6 %	84,4 %
ALD	3,9 %	3,9 %	92,2 %
HDD	1,3 %	2,6 %	96,1 %
ALR	-	1,3 %	97,4 %
OC	-	1,3 %	98,7 %
SB	-	1,3 %	100 %

We can see that the proposed frames make it possible to economically model a large proportion of the messages from each of the two trainee controllers, i.e. a limited number of frames describes a large number of exchanges. In addition, we can see that the model is constant between the two trainees. Their messages can be modelled using more or less the same frames.

6.5 Representing an aircraft

To complete the preliminary model of the dialogue, we must define the objects that the dialogue concerns, i.e. the status of an aircraft. We can visualize an aircraft as an object that moves following a trajectory through a space of states (i.e. the microworld of the aircraft), from the moment the trainee controller takes charge of it until it is handed off to another control centre. The states are the parameters covered by the messages exchanged between the pseudo-pilot and the trainee controller, as well as various other parameters describing the situation of a given aircraft. In fact, we are proposing here that the dialogue between the pseudo-pilot and the trainee controller can be considered a tool for moving the aircraft (or object) between different states. The following figure shows the various states of an aircraft that appears in one of the exercises, represented as an object.

Table VIII **Representation of the various states of an aircraft**

Model of an aircraft

Callsign: Lufthansa 471		
Previous sector:	Westport	
	Current	Next
Sector :	Eastview	Avalanche
Altitude:	FL 280	
Heading :	355	
Position:	15 m. S J565	
Altim.in.		
Frequenc:	133.5	132.35
Nor. nav:	N	
Follow up:	SIO Avalan.	

Note that the object describing the aircraft can also be considered a frame. Thus we can easily integrate this

representation of the aircraft into the language model. Within the frame, the various blocks constitute the states of the aircraft. In most cases, the current value of a status (marked "current" in the figure) comes from the instantiation of a sub-frame associated with it. The sub-frame corresponds to a part of the message received. The meaning of the different states is:

- Callsign: the aircraft's call sign, in this case Lufthansa 571. This status is fixed throughout the exercise.

- Previous sector: the sector from which the aircraft is coming, in this case Westport.

- Sector, altitude, heading, position, altimeter information, frequency: information on the current or upcoming ("next") situation of the aircraft.

- Normal navigation: the label is Y when the aircraft is handling its own navigation, and N when it is being guided by a controller.

- Follow up list: the stack of actions that must be taken. In this case, the aircraft will be coming under the control of a new controller (here Avalanche). It will have to report its presence, thus the SIO message waiting to be issued. When an aircraft receives an instruction, it must issue an acknowledgment. The acknowledgment remains on the stack until it has been issued.

6.6 Relations between the dialogue and the states of an aircraft

To determine the effect of the dialogue on the changing situation of an aircraft (or its representation in the state space), it is possible to relate the preliminary frames of the dialogue and the representation of the aircraft. That is done in the following figure.

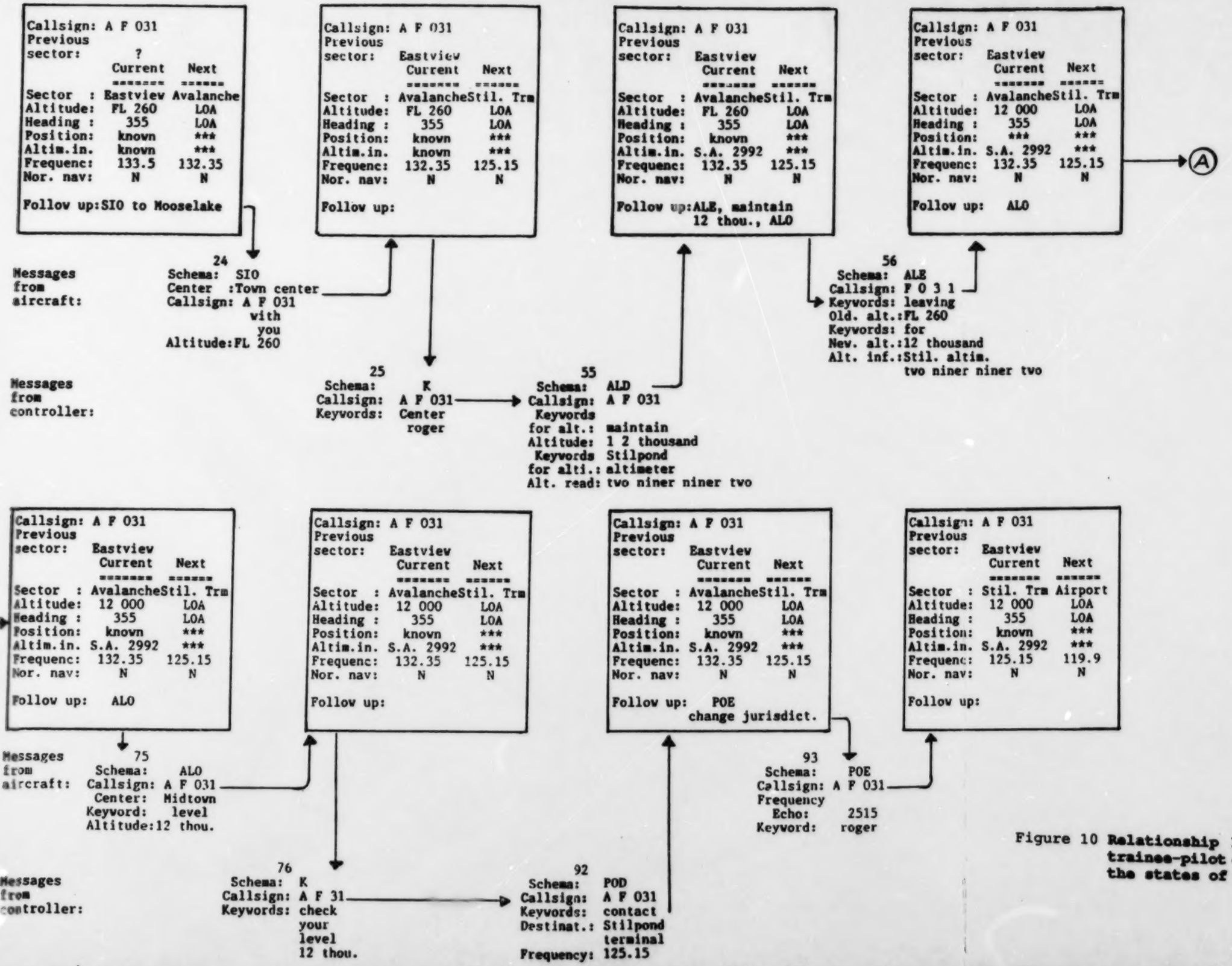


Figure 10 **Relationship between the trainee-pilot dialogue and the states of an aircraft**

The squares represent the "trajectory" of an aircraft in a space of states. The continuous line represents the sequence of states and transitions between states (caused by the messages exchanged). This gives us a first overview of the language model, where relations between the frames representing the operating language and the model of an aircraft are clearly shown.

In this chapter we developed part of the language model for simple exercises. The model is still incomplete, lacking mechanisms for the instantiation of frames and for error handling, as well as rules for calling successive frames. The next chapter will fill those gaps.

7. SECOND STEP IN DEVELOPING THE MODEL FOR SIMPLE EXERCISES

In this chapter we will complete the language model for simple exercises that was generally described in the previous chapter, by identifying the mechanisms for instantiating frames, handling dialogue errors and calling successive frames.

7.1 Rules for frame instantiation

In order to develop rules that can be used to help instantiate certain blocks in the frames, we first identify additional sources of available information during task execution, and how they can be used. We then examine the instantiation of various numeric blocks.

7.1.1 Identifying information sources

The fact that the task actually consists of a scenario and that the trainee controller works within a well-defined universe (i.e. the microworld of the controller) can work to our advantage. In addition, since the trainee can see an aircraft at any time on the radar screen, the air traffic simulator computer also has information on aircraft in the airspace.

The example in section 6.6 shows that only a few of the aircraft's states (e.g., altitude) are influenced by a given message. More generally, the following table shows the relationships between each of the frames obtained from the summary analysis and the status of the aircraft. There are a number of cases. First of all, a common element between a given frame (or message) and one of the states of the aircraft may simply be mentioned, and leads to no change in the microworld of the aircraft. This relation is identified by "A" in the table. A message received may refer to the aircraft's call sign, but since that state is constant, there is no change in the microworld of the aircraft. Then there is the case where a common element between the frame and the microworld of the aircraft changes the aircraft's microworld. There are two

possibilities. First, the change may be entirely predictable ahead of time (for instance, because of the exercise screnario) and calling the frame simply "confirms" the change. That is the case, in particular, for a hand-off to another control centre. The scenario for the exercise includes this change, and all the details are known either in advance (such as the frequency of the new control centre), or dynamically, by the simulation computer (such as the position of the aircraft). Those elements are identified by "B" in the table. Second, the change may be unpredictable. That is the case for a change in altitude, for example. The message must then provide all the information required. These elements are identified by "C" in the table. The assignment of ratings A, B and C is explained in Appendix F.

Table IX **Relations between the representation of an aircraft and frames**

Aircraft states	Frames HDR	HDD	VRD	VSD	ALDF	ALD	ALR	POD	K	OC	SB
Callsign	A	A	A	A	A	A	A	A	A	A	A
Previous sector											
Sector: current											
next								B			
Altitude: current											
next					C	C					
Heading: current											
next		C	C	C							
Position: current			B								
next											
Altimeter informat.: current											
next					B	B					
Frequency								B			
Norm. nav			B								
Follow-up	HDA	HDE	VRE	VSE	ALEF	ALE	ALA	POE			

The existence of alternative sources of information gives the possibility of compensating for errors introduced by the speech recognizer. Such a method must be used with care, however, as overuse might lead to correcting errors by the trainee.

7.1.2 Rules for instantiating certain sub-frames

In the previous chapter we presented a number of sub-frames containing numeric blocks ("altitude," "heading," etc.). Knowledge of the air traffic control situation, the status of a given aircraft, the history of the dialogue, etc., can be used to create rules to limit the values that these blocks may take on. The following table identifies some of those rules.

Table X **Limits for the instantiation of various sub-frames**

Sub-frame	Rule and justification
Next altitude	**minimum altitude (current sector) < next altitude = maximum altitude (current sector)** Justification : A controller cannot assign an altitude that is not part of the sector that he or she is controlling. Note : There are preferential altitudes for making hand-offs to other sectors. A more sophisticated model could take that into account.
Altitude to be reported	**if next altitude > current altitude, then current altitude > altitude to be reported > next altitude** **if next altitude < current altitude, then current altitude < altitude to be reported < next altitude** Justification : The "altitude to be reported" is crossed while going to the "next altitude".
Heading to reintercept	Rule : **The heading must make it possible to intercept the objective described in the message sent to the aircraft.**

This table is incomplete. Additional rules could probably be established for other (numeric or non-numeric) blocks, using the knowledge of an experienced controller.

7.2 Handling dialogue errors

This section examines how errors in the exercises are handled. To expand the data, we consider terminal control exercises in addition to low-altitude control exercises. The management processes used for handling errors are identified. We wish to determine rules likely to be used in dialogue between the trainee controller and the pseudo-pilot. By errors we mean events that occur in the dialogue and that are likely to mislead one or the other party. These are "serious" events (e.g., the new heading is misunderstood), as compared to other errors that may be tolerated by the frame-based model (e.g., an abbreviation of an aircraft's call sign).

7.2.1 Classifying error situations

An examination of the transcripts shows that there are various situations in which dialogue errors occur. The following table suggests a classification for these dialogue errors, in approximate order of complexity. The table gives all the situations where dialogue errors were observed in simple exercises (low-altitude control) and in complex exercises (terminal control).

Table XI **Classification of dialog errors**

Type of error	Messages relating to an error and issued by: trainee controller	Messages relating to an error and issued by: pseudo pilot	Total
Pilot omission and error	1 (0,5 %)	1 (0,4 %)	2 (0,4 %)
Error with self-correction in same message	12 (5,5 %)	-	12 (2,7 %)
Error with self-correction in following messages	1 (0,5 %)	1 (0,4 %)	2 (0,4 %)
Request for repetition	3 (1,4 %)	3 (1,3 %)	6 (1,3 %)
Identification and resolution of ambiguity	1 (0,5 %)	6 (2,6 %)	7 (1,6 %)
Total	18 (8,3 %)	11 (4,8 %)	29 (6,5 %)

The dialogue error situations may be described as follows:

. Omissions and errors: these errors occur when one speaker receives the correct instructions, but makes a "handling" error (e.g, the pilot takes the wrong

heading).

- Error with self-correction in the same message: these errors occur when a speaker commits an error, realizes it and corrects him- or herself in the same message.

- Error with self-correction in following messages: these errors occur when a speaker makes an error in a given message, realizes it once the message has ended, and sends another message to correct the error.

- Requests for repetition: these errors occur when one of the speakers asks the other to repeat some information.

- Identification and resolution of ambiguity: these errors occur when one of the speakers identifies an ambiguity and takes steps to resolve it. An ambiguity is not necessarily the direct result of an exchange of messages between the trainee controller and the pilot.

7.2.2 Error handling process

This section examines how the dialogue is managed, to identify and correct certain kinds of errors. Since our goal is to construct a dialogue model suited to an automated pseudo-pilot, we decided to eliminate cases in which the pseudo-pilot itself commits errors (pilot omissions and errors). We also eliminated cases in which the pseudo-pilot corrects itself in the messages it issues (the machine is not supposed to make any mistakes).

The error handling processes are analysed for the remaining cases in the same way as for the make-up of the frames. We first grouped together all error handling examples of the same type, and then tried to identify one or more patterns. These analyses are explained in detail in Appendix G. The following tables show the mechanisms observed for each case.

Table XII **Error-handling methods for errors with self-correction**

Error category	Mechanisms for handling observed errors
Self-correction within same message (by trainee controller)	ERRONEOUS SEMANTIC BLOCK + KEYWORD : CORRECTION + CORRECTED SEMANTIC BLOCK
	INCOMPLETE SEMANTIC BLOCK + KEYWORD : CORRECTION + CORRECTED SEMANTIC BLOCK
	INCOMPLETE ELEMENT OR WORD + KEYWORD : CORRECTION + CORRECTED SEMANTIC BLOCK
Self-correction in following messages (by trainee controller)	INITIAL ERRONEOUS MESSAGE + CORRECTED FOLLOWING MESSAGE CONTAINING THE KEYWORD "CORRECTION"

The rules required for handling an error with self-correction within the same message can be derived from the above table. They are:

. Rule: if the word CORRECTION is encountered in a message, the semantic block following that word replaces the similar block preceding the word CORRECTION.

. Rule: if two similar semantic blocks follow one another in the same message, the last semantic block will be considered valid.

For errors with self-correction in following messages, two cases must be considered. Either the automated pseudo-pilot detects the error (unlike the human pseudo-pilot, who occasionally cannot detect the error before the trainee does so) and an ambiguity identification and resolution mechanism is invoked; or the pseudo-pilot is unable to detect the error and, in that case, the contents of the following message containing the correction simply replace that of the erroneous message. Thus a single rule must be added to handle the latter case:

. Rule: if a message contains the keyword CORRECTION and is not a self-correction of that message, then the message is handled in the same way as a new instruction for the aircraft.

The following table deals with cases of requests for repetitions issued by the trainee controller.

Table XIII **Method of handling requests for repetition issued by the trainee controller**

Error category	Observed mechanisms for handling errors
Requests for repetition (by trainee controller)	RECIPIENT + KEYWORDS : "SAY AGAIN"

The following rule can be derived for handling requests for repetition issued by the trainee controller:

. Rule: if the message "Recipient SAY AGAIN" is issued by

the trainee controller, the last message from the aircraft must be repeated.

Then there are requests for repetition issued by the pseudo-pilot. The following table gives the results of the analysis of transcripts.

Table XIV **Method of handling requests for repetition issued by the pseudo-pilot**

Error category	Observed mechanisms for handling errors
Requests for repetition (by the pseudo-pilot)	The pseudo-pilot echos the message : Callsign + Valid semantic block (s) (optional) + KEYWORDS : "SAY AGAIN" + ambiguous semantic block

This situation can be modelled with the following rule:

- Rule: if the value of a semantic block is understood imperfectly or not at all, the pseudo-pilot can issue a request that the semantic block be repeated.

Then there is the most complex case of error handling, where one of the parties identifies an ambiguity and asks the other speaker to resolve it. The following table shows two types of mechanisms observed.

Table XV **Methods for identifying and resolving ambiguities**

Dialogue	Commentaires
MESSAGE ISSUED	The first party issues a message.
SEMANTIC BLOCK + "CONFIRM"	The second party replies by identifying an ambiguous element and requests confirmation.
"CONFIRM" OR "NEGATIVE" OR "CORRECTION" + SEMANTIC BLOCK	The first party confirms or corrects the ambiguous element. The error handling process is then complete.
MESSAGE ISSUED	The first party issues a message.
SEMANTIC BLOCK + "CONFIRM"	The second party replies by identifying an ambiguous element and requests confirmation.
The first party does not reply.	
REQUEST (NATURAL LANGUAGE)	After a moment, the first party asks the second party whether it has carried out the manoeuvre requested.
REPLY (NATURAL LANGUAGE)	The second party informs the first party that it has requested confirmation.
SEMANTIC BLOCK + "CONFIRM"	The first party confirms the element or not. The error handling process is then complete.

For the first case the error handling process is simple. The method can easily be modelled with the following rule:

- Rule: if the value of a semantic block is ambiguous, the recipient may formulate a request for confirmation of the semantic block in question.

It can also be seen that despite the various ways in which the recipient can formulate the confirmation, the rule for formalizing this mechanism is almost identical to that dealing with the requests for repetition from the pseudo-pilot.

The second method reflects an anomaly: an ambiguity has been identified, and the second party (in this case the pseudo-pilot) has tried to resolve it. The transcript shows that at this time the first party (in this case the trainee) has begun a conversation with another speaker (whether it is another aircraft or a control centre is not important). The request to resolve the ambiguity remained unresolved, and the pseudo-pilot did not repeat its request, instead taking no action. Finally, the trainee asked the pseudo-pilot for its intentions (the trainee either did not hear the pseudo-pilot's first request, or put it aside and forgot about it). The two speakers had to resort to natural language to pick up the dialogue where it was interrupted.

Although this situation seems more complex, it is probably possible to avoid the complications brought about by the hesitant recommencement of the conversation between the trainee and the pseudo-pilot. If the latter had reminded the trainee after the former had completed the conversation with another speaker, the error handling process would likely have been completed as in the first case. Thus we have provided the following procedure, to keep the error handling process in a well-defined framework:

- For as long as there is uncertainty about the value of a semantic block:
  - Wait for the conversation in progress between the trainee controller and another speaker to be completed.
  - Use the previous rule (the simple method for identifying and resolving ambiguities) to handle the situation.

Although this is a very primitive error handling process, it forces the speakers to handle ambiguities so that the pseudo-pilot can continue with its task. This insistence in the dialogue management process might have made it possible to resolve an incident wherein a trainee controller persisted in calling his interlocutor Nordair, whereas the aircraft had noted several times that its identification was Norontair.

7.3 Calling successive frames

In this section we will examine how the selection of frames to be used to classify a given message can be limited, in an effort to improve the performance of the speech recognition system.

7.3.1 Selecting the recipient for a message

The number of possible targets for a message can be limited. Only those aircraft in the control zone are likely to be called and to reply. If, by mistake, a message were sent to an aircraft not in the control zone (and which is not preparing to enter that zone), there should be no reply to the message, in that the aircraft is not on the radio channel used by the trainee controller. This gives us the following rule:

- Rule: only the aircraft present in the control zone or preparing to enter that zone may receive a message from the trainee controller.

7.3.2 Sequence of frames as the aircraft's status changes

When an aircraft is moving through a controlled sector, its trajectory and the messages it is likely to receive depend on a number of factors. For example, if it is moving through a low-altitude zone (in this case, Avalanche), messages relating to the final approach are not part of its microworld. Similarly, the allowable messages also depend on its position in the control zone. Thus an aircraft that has just arrived from another zone and that states its position in an appropriate message is not likely to receive a message handing it off to

another control zone for some time. Finally, and above all, it seems inevitable that the categories of messages received by an aircraft depend on the aircraft's scenario and the exercise in which it is involved. Thus we can state the following hypotheses:

- The order in which successive frames are called depends in part on the frames themselves: some frames simply do not occur in a specific part of a scenario, while others are almost inevitable.

- Different trainee controllers will use much the same frames to control aircraft in similar situations.

To check these hypotheses, we constructed the following table from the transcripts:

Table XVI **Calling successive frames**

Sequence of messages for the following aircrafts :									
N900PB		AC 951		LUF 471		AF 031		CAN 63	
Transcr.: 1a	2a	Transcr.: 1a	2a	Transcr.: 1a	2a	Transcr.: 1a	2a	Transcr: 1a	2a
				SIO	SIO	SIO	SIO	SIO	SIO
				K		K	K	K	SB
HDR				HDR	HDR	ALD	ALD		
HDA				HDA	HDA	ALE	ALE	( )	
VSD				VSD	HDD	ALO	K		ALR
( )				( )	HDE	K			ALA
ALDF	ALDF	ALD	ALD	ALDF	ALDF				K
ALEF	ALEF	ALE	ALE	ALEF	ALEF			ALD	ALD
ALF	ALF			ALF				ALE	ALE
	K			K					
VRD	VRD			VRD	VRD				
VRE	VRE			VRE	VRE				
[VRD]				ALO	ALF				
K				( )					
				ALO					
				K					
POD	POD	POD	POD	POD	POD	POD	POD		POD
POE	POE	POE	POE	POE	POE	POE	POE		POE
			K						

First of all, the table shows the frames exchanged between the trainee controller and the pilot, for a number of aircraft. We have included the frames modelling the pilot's replies, i.e. SIO, HDA, ALEF, ALF, ALO and VRE, although we will not consider them in this analysis. Finally, it can be seen that part of the sequence of frames for certain aircraft appears to be missing; that is because the recordings of the exercises started at different times, while the exercises were already in progress.

This table seems to confirm the first hypothesis, that some frames apparently occur only at the beginning or end of the dialogue between each trainee controller and each pilot (e.g., SIO, POD, POE). That is in fact an obvious conclusion, given the very nature of the frames.

In addition, there is a very strong correlation between the frames called for a single aircraft controlled by two different trainee controllers. They use essentially the same frames, in much the same order, which also tends to confirm the second hypothesis.

An examination of the transcripts also confirms that not all the aircraft need all the available frames. For example, AF 031 and CAN 63 use only part of the frames used by LUF 471 and N900PB. It also seems obvious that barring an error by the trainee controller, which is always possible, only the pertinent sub-set of all the frames available will be used to control the aircraft. Thus we can say that there is a set of frames available to carry out a given control task, and that set consists of "sub-sequences" of the frames necessary to perform the task. For example, handing off an aircraft to another zone calls for the sequence POD-POE, possibly followed by a frame K.

We will now propose specific rules concerning some frames, as well as more general rules concerning the use of a set of frames for a given aircraft:

- Rule: when an aircraft arrives in a control zone, it generally sends an SIO message. The trainee controller may reply with a K, SB or another message.

- Rule: if an aircraft is not at the preferential altitude for entering the next sector (i.e. its destination), a sub-sequence for changing altitude shall be used at the proper time prior to a POD message.

- Rule: if an aircraft follows a heading different from the preferential entry heading for the next sector, a sub-sequence for changing the heading shall be used at the proper time, prior to a POD message.

- Rule: the last sub-sequence in controlling an aircraft is a sub-sequence comprising POD-POE, possibly followed by a K message.

7.4 Resulting model

This section presents part of the language model that has been developed thus far. We now have:

- a skeleton model consisting of blocks that illustrates the architecture of a dialogue management system, and incorporates part of the error handling mechanisms;
- frames that integrate the other error handling mechanisms and the various other changes noted thus far.

It should be emphasized here that the proposed model is only one possible conceptualization of the architecture of the dialogue model. To incorporate the improvements identified in this chapter, we had to decide on a number of items that had previously remained at the concept stage. Thus we put those concepts into practice in a given form, and we do not claim that that form is unique, or the best. It simply allows us to combine the elements obtained into a coherent and potentially usable whole. The following figure illustrates the skeleton of blocks that represents the architecture of the dialogue management system.

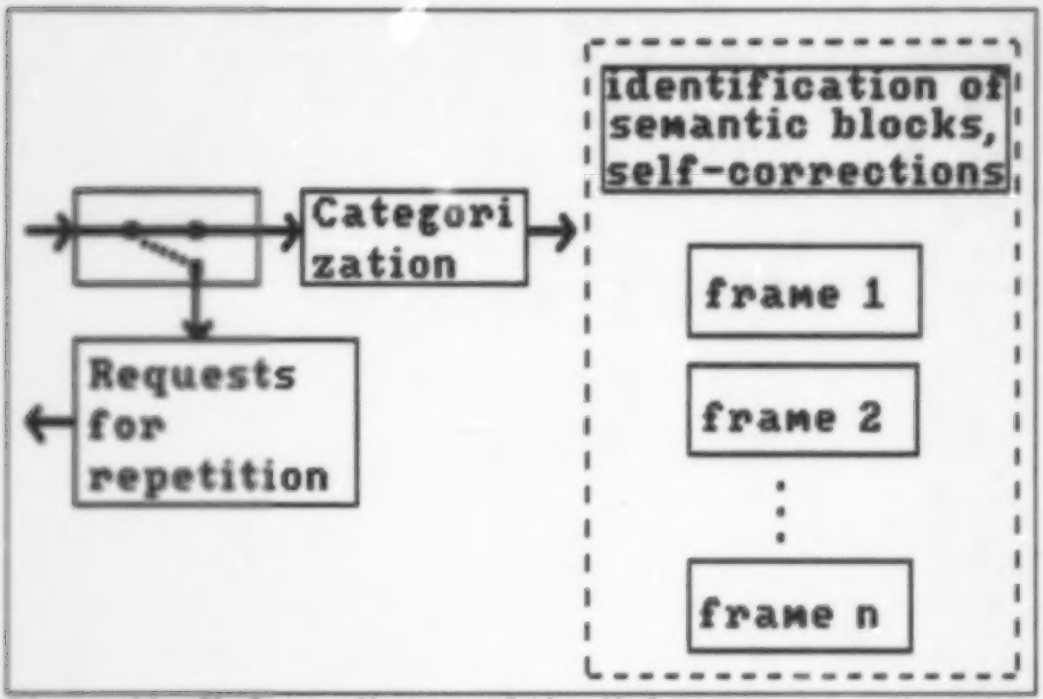


Figure 11 Skeleton diagram of the dialogue management system

It can be seen that rules for error handling have been distributed throughout the system. This distribution corresponds functionally to the mechanisms observed:

- Requests for repetition from the trainee controller are handled with no attempt to identify the subject of the semantic block repeated.
- Self-corrections seem to be handled during the instantiation of the frame identified.
- The identification and resolution of ambiguities are handled in the blocks concerned, for the frame to which the message belongs.

The figure below shows an example of a frame incorporating some of the improvements described in this chapter:

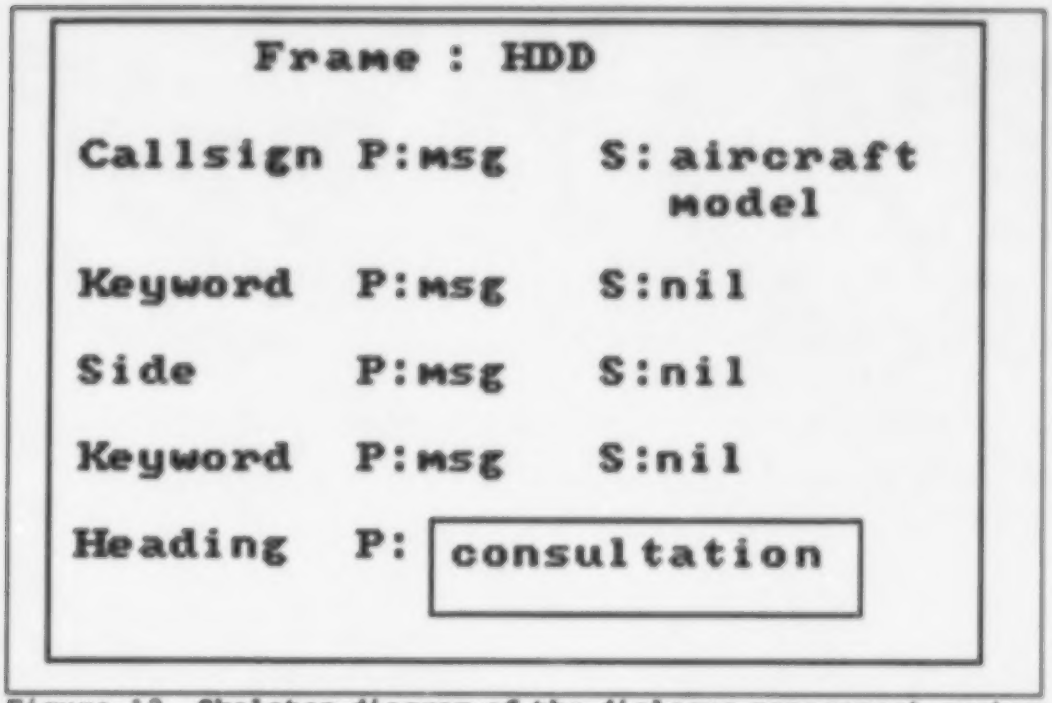


Figure 12 Skeleton diagram of the dialogue management system

For each of the slots in the frame, the letter P means the main source of information, and the letter S means a secondary source. For example, for the Callsign block, if the corresponding part of the message is defective, the aircraft model for which the message is intended can be used to make a correction. The following blocks (Keyword, Side, Keyword) contain keywords or values that have no essential semantic content for the moment. The last block (Heading) illustrates the last error handling mechanism: identification and resolution of ambiguity. The transcripts show that for an ambiguity to be identified, it must be significant, and the pseudo-pilot must be able to detect it. For example, an error concerning one of the letters of the call sign does not generally justify the process. An ambiguity concerning vital information, however, is detected and must be resolved. The transcripts show that this vital information is altitude and heading, in the exercises examined. Unfortunately, the only source for that information is the message itself. The message may be correct, or may be faulty owing to an error by the speech recognizer or the trainee controller. Thus some of the rules

suggested for instantiating certain sub-frames may be used to validate and, if necessary, correct those blocks. That is the purpose of the "consultation" process shown in the above example. That process receives the value extracted from the message for the block concerned and checks whether that value is plausible, in view of the instantiation rules for this sub-frame (altitudes and preferential routes, probability of the message given the aircraft's current status and the history of the dialogue, etc.). The consultation process may be considered, for example, a backward-chaining expert system: starting with a conclusion (i.e. a proposed altitude or heading), an attempt is made to determine whether the conditions leading to that conclusion are present. If applicable, the consultation process indicates that the proposed conclusion is plausible. If not, the process indicates that the conclusion is implausible, and may propose a different conclusion, plausible in the circumstances.

The model presented here describes the language used in simple exercises. The following chapter looks at whether that model can be used in a more complex situation, i.e. terminal control. We will then have taken an overall look at the modelling of the language used by trainee air traffic controllers.

8. EXTENSION OF THE RESULTING MODEL

In the two previous chapters we designed a dialogue model making it possible to understand conversations between a trainee controller and a pilot in a simple exercise.

We now wish to determine whether this simple model can be extended, i.e. whether it can be adapted to a more complex situation. As mentioned in Chapter 3, the most complex situation is terminal air traffic control. Accordingly, we made two recordings of a terminal control exercise at the Transport Canada Training Institute in Cornwall. Each 45-minute recording corresponds to an exercise by a different trainee controller. The transcripts are referred to as 3a and 4a, and are reproduced in Appendix H.

We also wish to determine how to relax the constraints on the concatenation of messages within a transmission and on the number of recipients to whom a transmission is sent.

8.1 Adaptability of the model

To find out whether the simple model can be adapted to terminal control, we must:

- determine the approximate size and composition of the operative vocabulary for terminal control operations;
- determine whether the frames of the simple model can be used to analyse part of the conversations between the trainee controller and the pseudo-pilot in terminal control operations;
- determine whether the same analysis method can be used as for the simple model, to add the frames necessary to complete the model.

8.1.1 Length of transmissions in a terminal control situation

As with the simple exercises, we wish to evaluate the length of transmissions from the trainee controller. The following figure first shows the distribution of the length of transmissions in a terminal control situation:

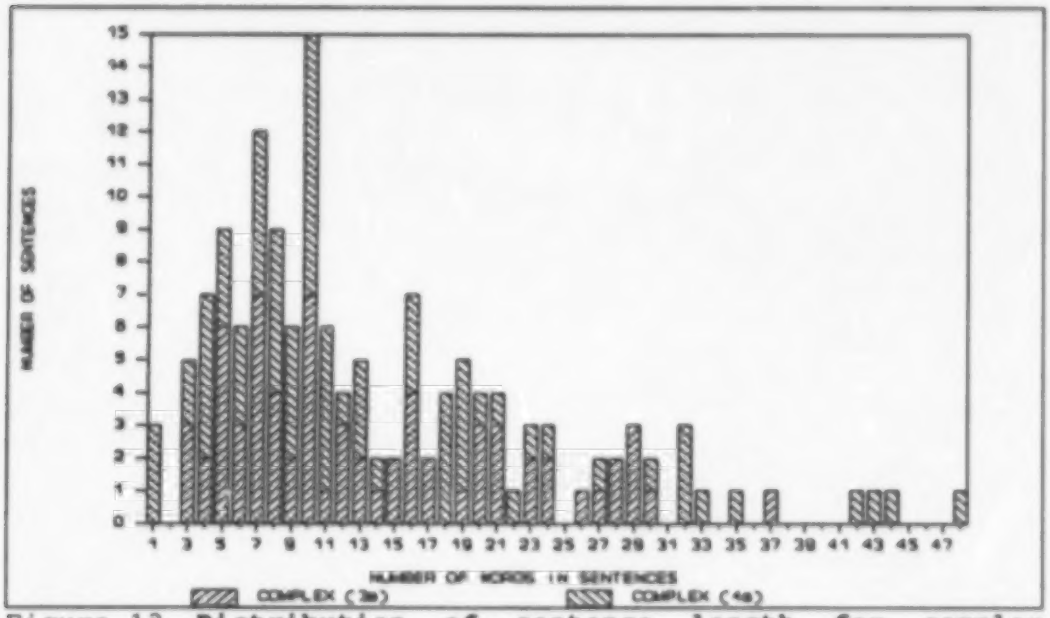


Figure 13 **Distribution of sentence length for complex exercises**

Now we wish to determine how the distribution of sentence lengths for the complex exercises compares with that in the simple exercises. That comparison is given in the following figure.

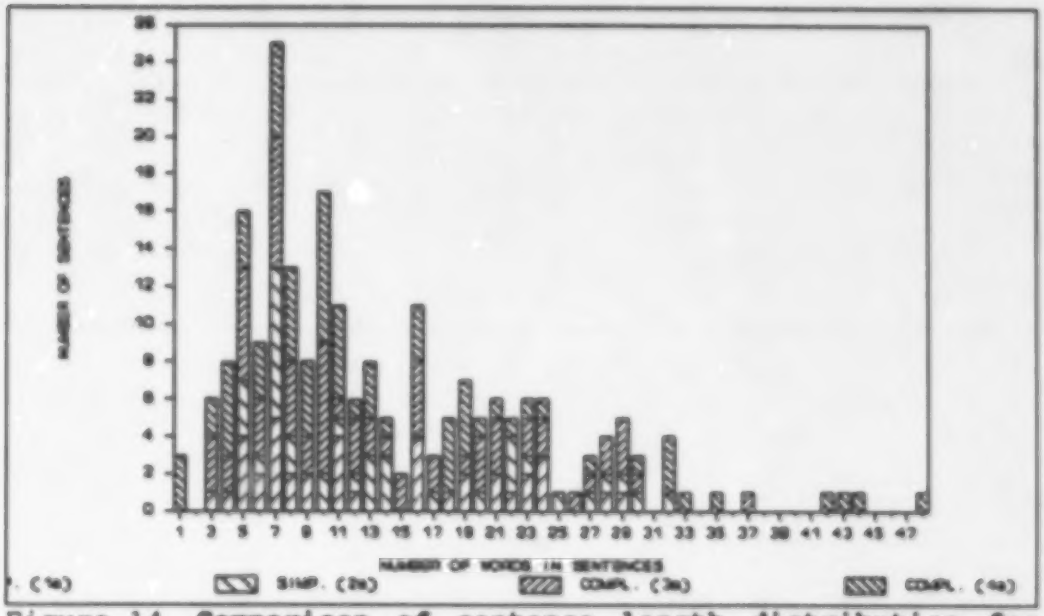


Figure 14 **Comparison of sentence length distribution for simple and complex exercises**

The figure shows that in the simple and complex exercises examined, the length of most of the transmissions varies from 3 to 30 elements.

8.1.2 Operative vocabulary for terminal control

In this section we evaluate the size of the vocabulary required for terminal control tasks. We proceed in the same way as for the simple model, by determining the net operative vocabulary (incomplete) for each transcript. We then find the combined net operative vocabulary (incomplete) for the two transcripts. Finally, we add a standard number of words to take account of numbers, names of control centres and airlines, etc. The following table gives the size of the incomplete net operative vocabulary (the list of words in the vocabulary is given in Appendix I) for those exercises.

Table XVII **Size of the incomplete net operative vocabulary for terminal control exercises**

	Transcript 3a	Transcript 4a
Net operative vocabulary (less references to numbers, name of airlines, etc.)	60 words	71 words

First of all, it can be seen that the vocabularies for the two transcripts are roughly the same size. As with the simple model, we determine the size of the necessary vocabulary for the two transcripts together, and obtain a net combined vocabulary of 81 words. Thus it appears that different trainee controllers use more or less the same vocabulary. We then add the standard number of required words (numbers, control centres, etc.) and obtain:

Table XVIII **Size of the operative vocabulary for terminal control exercises**

Size of the incomplete net operative vocabulary	Number of standard elements	Size of total operative vocabulary
81 words	79 words	160 words

If we add a safety margin of 25%, we obtain a total operative vocabulary of (160 x 1.25) 200 words for terminal control operations.

We conducted an additional analysis to determine the similarity of the vocabulary for simple and complex tasks. We considered only the net operative vocabulary, since the number of standard elements is the same for the two types of tasks. The incomplete net operative vocabulary making it possible to describe the two tasks simultaneously is 85 words, e.g. with the exception of four words, the entire vocabulary for the simple task is included in the vocabulary for the complex task. This is an interesting point, as it gives us the following choices:

- Use a different vocabulary for each type of task: we can then use a small vocabulary for simple tasks, possibly making for better performance by the speech recognizer.

- Use a single vocabulary comprising the simple and complex tasks, and thereby simplify the system.

8.1.3 Usefulness of simple model frames for terminal control operations

We can evaluate the usefulness of the frames for the simple model as components of the dialogue model for terminal control operations by examining what proportion of the messages between the trainee controller and the pilot can be "explained" or analysed with the frames of the simple model. If we find that many of the frames are "re-usable" in the terminal control situation, it suggests that the simple model is robust and can possibly be expanded. If, on the other hand, we find that few of the simple model frames recur in the control terminal dialogue, it suggests that we need a different model for each control sector. If that is the case, this model would be difficult to use in practice.

We begin by extracting the messages issued by the trainee controller from the terminal control exercise transcripts, and then attempt to use the frames defined in the preliminary model to describe those messages. The results are shown in the following table.

Table XIX **Usefulness of simple model frames for terminal control situations**

Situation	Transcript 3a	Transcript 4a
Case where simple model frames are sufficient for analysing messages between the trainee controller and the pseudo-pilot	55.7 %	59.4 %
Case where simple model frames must be expanded in order to be able to analyse messages	8.6 %	10.8 %
Case where supplementary frames are required in order to analyse conversations between the trainee controller and the pseudo-pilot	35.7 %	29.7

The table shows that almost two-thirds of the conversations between the trainee controller and the pseudo-pilot can be described using either the original frames devised for the simple model, or the original frames with some minor changes. This observation is true for each of the terminal control transcripts (different trainee controllers, same task). The most important consequence of this observation is that the approach used thus far to model the language (i.e. frames) and the method used to devise the frames appear to be appropriate.

This observation is all the more significant in that, as mentioned in Chapter 6, the samples are very small, indicating that the dialogues are stable. When the table was being prepared, we also noted that:

- The frames that will have to be added for terminal control operations are more or less identical for the two transcripts, confirming that the dialogue is stable for different trainee controllers.

- There is a relatively small number of new frames to add. At first glance, it appears that five new frames should be sufficient to complete a satisfactory model for the terminal control exercises observed.

We made no changes to the original frames, since they are likely to change, as we mentioned earlier, when the model is implemented.

Appendix J also provides draft additional frames necessary to model the terminal exercises (for the trainee controller and the pseudo-pilot).

Finally, we note that additional frames can be defined here in a number of ways: there is more than one way of dividing a frame. For example, one of the new frames drafted (CFAD) could be either a new, independent frame, or a new frame using some of the frames from the preliminary model (in this case, ALD and HDD). It appears that an iterative approach using the additional draft frames could be used during a simulation. The results of the simulation and other results on the optimization of frames (29) could then be used to refine the model.

8.2 Relaxing constraints on the concatenation of messages and the number of recipients of a message

Thus far we have failed to address a complexity found in the transcripts: some transmissions carry more than one message for the same recipient, and some carry different messages for more than one recipient. This concerns only transmissions from trainees, since those from pseudo-pilots involve no modelling problems. The following sections summarize the results of the analysis of this type of messages. The detailed analysis is given in Appendix K.

8.2.1 Several messages for the same recipient

This table summarizes the analysis for the case of several messages for one recipient.

Table XX **Analysis of transmissions containing several messages for one recipient**

Callsign 1	+	message K	+	message 2
Callsign 1	+	message 1	+	message 2
Callsign 1	+	(message 1)	+	message 2

Three main patterns can be seen:

- a transmission comprises the recipient's call sign, a category K message and a second message;
- a transmission comprises the recipient's call sign, a message of some category, and a second message;
- a transmission comprises the recipient's call sign, a message of some category, and a second message. An error handling process may also be included in the transmission (for instance, self-correction of one of the messages in

the transmission);

- finally, we found a maximum of two different messages per transmission. Nonetheless, a larger sample might show that it is possible for a transmission to include a greater number of messages.

This situation is modelled with the following rules:

- Rule: a transmission is sent to a single recipient, and comprises more than one message category if a) there is only one call sign in the message (apart from that of the person issuing the message) b) a single message category does not satisfactorily explain the contents of the transmission.

- Rule: the reply to a message issued by an aircraft may be a transmission beginning with a message corresponding to frame "K" (frame K represents an acknowledgment) followed by another message for the same aircraft.

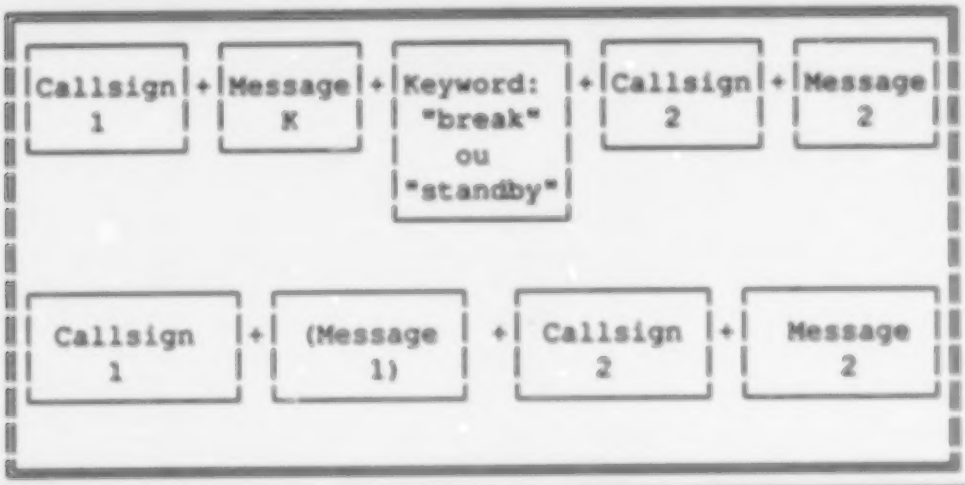
- Rule: a transmission including more than one message and sent to a single recipient may also contain one or more error handling processes.

These rules are general enough, since they do not attempt to define the type of messages that may be concatenated in a transmission: that would most likely call for a knowledge of the details of the exercise, as well as in-depth knowledge of air traffic control.

8.2.2 Different messages for more than one recipient

The following table illustrates the analysis of the remaining case.

Table XXI **Analysis of transmissions containing different messages for more than one recipient**



An analysis of the transcripts shows that there is a maximum of two recipients for a transmission. Even though a given transmission could concern more than two recipients, that seems highly improbable. Indeed, the possibility of errors on the part of the trainee controller seems too high (a great deal of information has to be sent in a short time) and dialogue management becomes very difficult (each instruction involves a confirmation message from each aircraft). Thus we consider that a transmission concerns at most two recipients. The situation is modelled with:

- Rule: a transmission includes more than one message and is intended for more than one recipient if a) the "Callsign" element is repeated, followed by "message," for different call signs, b) the keywords "break" or "standby" are included (optional).

- Rule: there are at most two recipients for a transmission.

As soon as the "Callsign" element followed by "message" have been recognized, each part of the message can be analyzed

separately.

8.3 Comments on the modelling of the language

Thus far, we have shown that:

- It is possible to build a frame-based language model for a simple air traffic control exercise.
- The elements of the simple model (vocabulary, frames, etc.) can be used to build a language model for a complex exercise.

Since we have not considered all types of exercises or all the air traffic control situations used at the Training Institute, the model developed thus far is not exhaustive. However, it can be used as the basis for developing an automated pseudo-pilot system. The language model can then be improved as experience with the system grows. It seems clear that improving the model will call for the co-operation of a person knowledgeable in the field of air traffic control. This concludes the discussion of the modelling of the language.

The last objective in this study is to examine the use of speech recognition and generation technologies for use in an automated pseudo-pilot. That is done in the following chapters.

9. REQUIREMENTS AND POSSIBILITIES FOR SPEECH RECOGNITION

In this section we will examine several aspects of speech recognition as it applies to automated pseudo-pilots. We begin by giving a short description of the type of speech recognizer required; we then examine a number of ergonomic factors such as the training process and the effect on dialogue of interacting with a machine.

9.1 Required speech recognizer features

If we wish to obtain an automated pseudo-pilot with vocal behaviour similar to a human's, i.e. one that is transparent for the trainee controller, it must have the following features:

- connected word recognition, as in the case of the existing pseudo-pilot;
- reaction time (including the time required by the speech recognizer) comparable with that of a human pseudo-pilot, which has been evaluated at roughly 1.75 seconds;[5]
- activated by a foot-operated switch, as in the existing workstation;
- vocabulary of about 200 words; this number corresponds to the total vocabulary for all the exercises examined here. We can still use a smaller vocabulary for certain exercises if a vocabulary adjusted to each exercise is employed;
- depending on the strategy adopted concerning the syntax (strict, evolutive or flexible), we must provide

[5] This value corresponds to the average time between the sending of the message by the trainee controller and the reply by the pseudo-pilot. We obtained this value by taking four measurements, at random, for each of the four recordings used in this study. The values obtained (rounded off to the nearest second) for the recordings are 3, 2, 1, 3, 1, 1, 1, 1, 1, 1, 2, 3, 1, 1, 2 and 4, respectively. Thus the average time is 1.75 seconds.

different abilities in terms of the number of branchings allowed for the finite-state speech recognizer. For flexible syntax, this is not a problem. For evolutive syntax, we must be sure that the speech recognizer is designed to switch quickly among various syntaxes;

- able to analyse a sentence comprising from 32 words (for a simple exercise) to 48 words (for a complex exercise).

The system must also be relatively insensitive to conversations taking place at adjacent workstations, which suggests the use of a directional noise-cancelling microphone. Given the behaviour of certain speech recognizers, the same microphone should be used for training and operatiing the speech recognizer. In addition, directional microphones are sensitive to their orientation, by definition, which can result in appreciable variations in the signal level (30 times). We still do not know whether this will be a problem in the automated pseudo-pilot, but if so there are some possible solutions:

- Wearing of headsets: we can probably encourage trainee controllers to wear their headsets correctly and in a standard fashion during the training on the simulator.

- Level of elocution: we can first of all encourage clear and distinct elocution by trainee controllers, for example during a familiarization session with the speech recognizer. During regular operations, we can rely on an additional method, displaying the sound level received by the speech recognition module on an indicator panel, and allowing trainee controllers to adjust their elocution accordingly. This latter method has the inconvenience of distracting the trainee to some extent, simply to assist the speech recognition system, and so does not appear to be a very acceptable solution (the trainee controller's task is to learn about air traffic control, not to make him- or herself heard by the machine).

- Input gain of the speech recognition module: some speech recognizers include automatic gain control of the input signal (ITT VRS 1280, SSI Phonetic Engine, Marconi ASR 1000, etc.). The gain is typically adjusted after an entire sentence has been entered. This method (combined with correct wearing of the headset) appears to be the least inconvenient for trainee controllers.

It would be ideal to use a speaker-independent speech recognition module. However, we do not have sufficient information on connected word speaker-independent speech recognizers, for the vocabulary size under consideration, to determine whether that would be feasible. We therefore suggest the use of a single-speaker speech recognizer, which we feel would be likely to provide the best recognition rate. That requires that we train the speech recognizer by providing voice samples for each user. Thus it is important that the system chosen have user-friendly tools for obtaining speech samples. To improve the quality of the samples, it is also important that multi-pass training be possible.

9.2 Training with the speech recognizer

One of the factors that seems most important in the successful application of speech recognition is the user's attitude (31) (32). Most users come into contact with the speech recognizer for the first time during the training phase. Thus it is very important that the training phase be used to encourage co-operation by the user, as well as to obtain high-quality voice samples, obviously, for the speech recognizer. We propose a training framework that could be used to attain these goals:

- First, a phase during which the trainee controller familiarizes him- or herself with the speech recognizer, with a limited vocabulary. The best results in recognition are generally obtained for users who have had the time to familiarize themselves with the speech recognizer (31). In addition, various experiments suggest that it is best to proceed progressively where the vocabulary is concerned. This initial training phase could be held during a session when the trainee controller becomes familiar with certain parts of the phraseology used in air traffic control, such as the ICAO radio-telegraphy phonetic alphabet (alpha...zulu), numbers, and so on. This is in isolated-word mode.

- Then an intermediate phase, where the vocabulary is completed and connected speech is introduced.

- Finally a training phase for phraseology in connected speech mode.

Other scenarios could easily be devised; the important point is that training be considered an opportunity to be exploited, rather than an obstacle to be overcome. One interesting possibility would be to integrate training of the speech recognizer with training on certain parts of the phraseology. That could be useful for the training on phraseology in general, and for the training of the speech recognizer in particular. In fact, this point was raised during discussions with pseudo-pilot systems designers (9): they felt it could be profitable to familiarize trainee controllers with speech technology through a simple system (e.g., a "game", where the trainee would guide a light symbol along a given path, using voice commands).

There are other considerations relating to the training phase, and which can contribute to improved performance:

- The best results are obtained when training is supervised, and carried out in an acoustic environment as close as possible to the operating environment (32) (33).

- Training should be conducted in such a way as to reproduce the tone of voice that will be used by the trainee controller while actually working with the speech recognizer; a user articulates differently depending on whether he or she is reading a text or giving commands to a system (34).

9.3 Effects of interaction with a machine on language behaviour

We know that users do not speak in the same way when they think they are talking to a machine, rather than a human. In a study on man-machine dialogue (35), it was observed that interacting with a machine can cause the following behaviours among humans:

- The dialogue is stabilized, i.e. communications tend to be close to a "standard" form.

- The dialogue tends to be less interactive.

- The user tends to imitate the tone and the slow and

monotonous rhythm of the machine.

It is difficult to determine the exact cause (fear of not being understood, effect of a synthetic voice, etc.) of this behaviour; the fact remains that it lends itself well to speech recognition. Even though the study in question was carried out in another context (requests for train schedules), it is reasonable to believe that the same observations could apply in the case of an automated pseudo-pilot. Thus it could be profitable to try to encourage this "stabilized" language behaviour, for example by properly explaining an automated pseudo-pilot's limited ability to understand.

Without claiming to be exhaustive, this chapter has emphasized various aspects of speech recognition that must be taken into account to simplify the operation of an automated pseudo-pilot. The next chapter examines speech generation, and suggests a choice suited to the task, the technologies available and ergonomic factors.

10. REQUIREMENTS AND POSSIBILITIES FOR SPEECH GENERATION

In this section we will attempt to identify a speech generation method suited to an automated pseudo-pilot system. We first examine the pertinent features specific to the task. We then briefly describe two of the most common speech generation methods. Next, we examine the ergonomic aspects of the synthetic voice, from two points of view: its perception, and the effects of its use on the language behaviour of the user. Finally, we select a method, taking account of the task, speech generation technologies, and ergonomic aspects.

10.1 Pertinent features specific to the task

The first observation that can be made upon examining the transcripts is that trainee controllers at the Transport Canada Training Institute at Cornwall generally deal with the same individual throughout a given exercise. The latter plays the roles of the different pilots, as well as the controllers of other control zones. Thus only one voice is used per trainee controller and per exercise in the present system.

As was shown previously (Chapter 6), messages issued by pseudo-pilots are fairly stereotypical, although less so than messages from trainees. There are a number of variations, relating mostly to the different abbreviations used by the different pseudo-pilots (the pseudo-pilots may change from one exercise to another) and their use of greetings (e.g., "good day" or "good afternoon"). We have already prepared the frames (see Appendix D) corresponding to the messages issued by the pseudo-pilots during the development of the preliminary dialogue model.

We then determine the size and composition of the vocabulary necessary for the pseudo-pilot's task, using the same method as with trainees' vocabulary. The following table shows the results of this analysis for all the transcripts (i.e. simple and complex exercises) used thus far. The complete list of words is given in Appendix L.

Table XXII **Vocabulary used by pseudo-pilots in the four transcripts**

Simple exercise (low altitude)
(Transcripts 1a and 2a combined)

Size of general vocabulary	Number of standard elements	Size of total vocabulary
56 words	79 words	35 words

Complex exercise (terminal)
(Transcripts 3a et 4a combined)

Size of general vocabulary	Number of standard elements	Size of total vocabulary
86 words	79 words	65 words

If a safety margin of 25% is added (as for the vocabulary of trainee controllers), the vocabularies required by pseudo-pilots during simple and complex exercises come to 169 and 206 words, respectively (i.e. 135 x 1.25 and 165 x 1.25). In addition, as with the trainees, a detailed look at the analyses shows that the pseudo-pilots used essentially the same vocabulary, although the vocabulary used by pseudo-pilots is more variable than that used by trainees.

10.2 Description and characteristics of generation methods

We will now very briefly describe the two most common methods of automatic speech generation: concatenation of words or sentence segments, and text-to-speech synthesis. We wish to identify the pertinent functional characteristics of these methods for purposes of their application in an automated pseudo-pilot.

10.2.1 Description and characteristics of word or sentence segment concatenation

This is done by recording the elements (words or sentence segments) of the messages to be generated and reconstituting them on request by concatenating them in a message (36). In this way we can create a message from parts, word by word, or use a set sentence to which variable elements are added (like a "speaking clock"). The following figure shows the main steps used to obtain the words or sentence segments that will be used for concatenation:

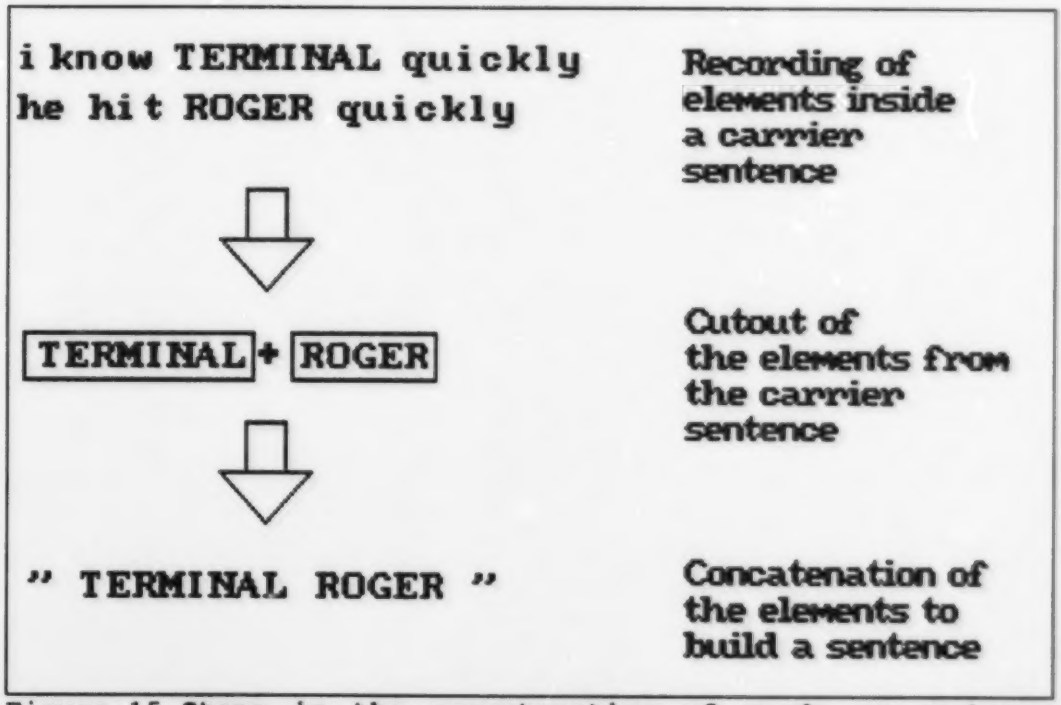

i know TERMINAL quickly
he hit ROGER quickly
Recording of elements inside a carrier sentence
TERMINAL + ROGER
Cutout of the elements from the carrier sentence
" TERMINAL ROGER "
Concatenation of the elements to build a sentence

Figure 15 **Steps in the concatenation of words or sentence fragments**

To use this method, we first of all record each of the words or sentence segments (i.e. a coherent set of words) from a "carrier sentence," in order to obtain a neutral sample (i.e. with no particular intonation) and one with no co-articulation effect with other elements. The sample obtained is then digitized and stored either in raw form (i.e. without

compression), or in coded form (i.e. with compression). Compression is intended to reduce the space required for storing the samples; there are standardized coding algorithms that can be used to reduce storage needs substantially with no appreciable deterioration in speech quality (37). The importance of compression can be better illustrated with an example. If we assume a bandwidth of 5 kHz for the voice to be used, and digitize that voice with a 12-bit analog-to-digital converter, we obtain:

5 kHz x 2 samples/cycle x 12 bits/sample = 120,000 bits/sec

That corresponds to 15 kbytes (120,000 bits/(8 bits/byte)) of computer storage space for each second of speech recorded. At that rate, each minute of speech recorded represents nearly one megabyte of storage space. In comparison, coding the digitized signal using an ADPCM-32 kbits/sec algorithm reduces storage needs to 4 kbytes per second of recorded speech, for nearly a fourfold savings in storage space.

For word or sentence segment concatenation, the intelligibility of the resulting voice is generally good (38). On the other hand, since the words or sentence segments are neutral, the concatenated voice is prosodically quite different from a natural voice. Research has been done in an effort to improve this aspect of the method (39) (40), and there are now ways of generating a concatenated voice with certain prosodic features. However, we lack evaluations of the impact of such methods on voice quality. Those modifications also require additional manipulation to generate the concatenated voice.

Then there is also the aspect of the flexibility of word or sentence segment concatenation. Unlike other methods where a word is built from smaller units (phonemes, diphones, etc.), in this case we can concatenate only the words or sentence segments already recorded. An extension of the vocabulary requires that words or sentence segments, preferably using the same speaker (it would be odd to hear different voices in the same sentence), be added to the units already stored. Since each word or sentence segment must be stored as a separate unit, this method is clearly suited only to limited and stable vocabularies (i.e., unlikely to change much).

10.2.2 Description of text-to-speech synthesis

As its name suggests, text-to-speech synthesis makes it possible to enter words and sentences in a system and to obtain a synthetic voice pronouncing those words or sentences. The following figure illustrates the main components of a complete text-to-speech synthesis system (adapted from (41)).

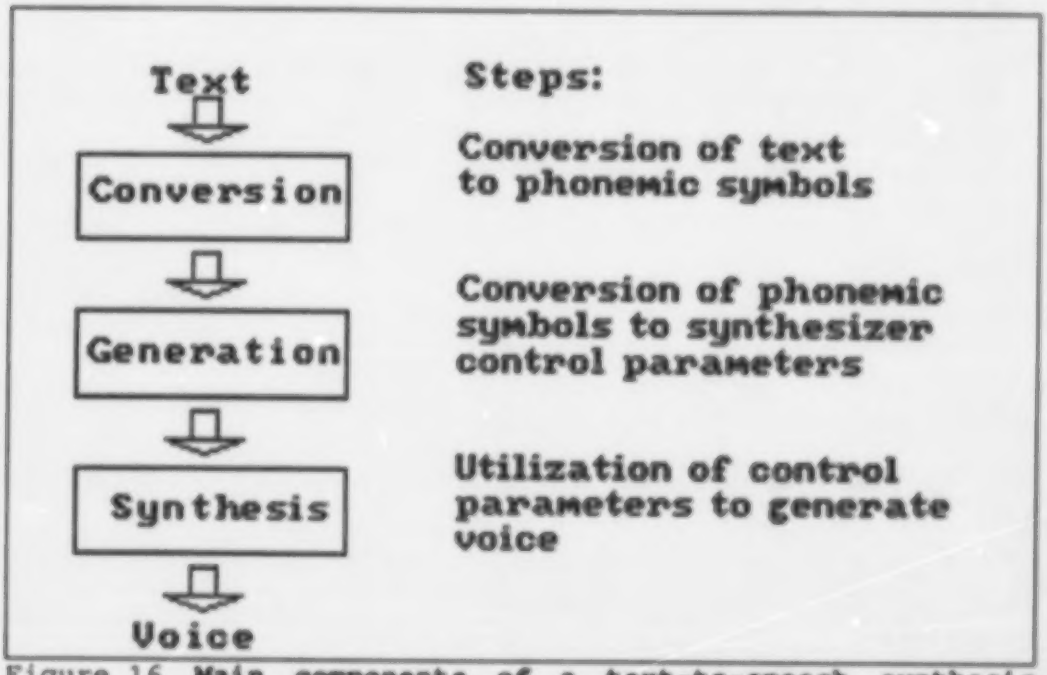


Figure 16 **Main components of a text-to-speech synthesis system**

Since these systems are required to handle any and all words or sentences, it is not feasible to record all the words likely to be encountered. We begin by "translating" words into "basic" acoustic elements smaller than words, i.e. phonemes, to generate a synthetic voice. A phoneme is the smallest individual sound in a language. There are a limited number of phonemes in each language; in English, for example, there are about 40.

To generate the synthetic speech, we must manipulate or assemble the phonemes satisfactorily. The idea of concatenating the phonemes making up each word is attractive, but there is a major obstacle: it is not the phonemes themselves that hold information in speech, but rather the transitions between

phonemes. Pronunciation of a sound depends on the preceding (pronounced) sounds and the upcoming sounds (that the speaker is preparing to pronounce), among other factors. Pronunciation is also affected by intonation and flow.

Many approaches can be used to assemble the phonemes, taking account of transitions among them. The most popular method is a rule-based approach to modelling those transitions, and its features are:

- uses a minimal amount of memory to store all the components (phonemes) of the sounds in a language;
- requires a large number of rules to represent the transitions between phonemes. The quality of the synthesis depends on the form and number of rules in the system.

Since phonemes vary from one language to another, a rule-based synthesis system works best in the language for which it was developed.

10.3 Ergonomic aspects: perception of generated speech

In this section we will examine various ergonomic aspects that may influence the choice between the speech generation methods described above. In particular, we will look at:

- the "performance" of various synthetic voices: this aspect takes in the traditional indicators used in evaluating synthetic voices, such as clarity, intelligibility, naturalness, acceptability, etc.;
- the cognitive load of processing a synthetic voice, as compared with a natural voice.

These two aspects are interesting mainly because of their direct link with the task of the trainee controller. The intelligibility of certain critical information (for instance, altitude, heading, etc.) is of vital importance. Similarly, we also wish to use a voice that is close to the voice of a human pseudo-pilot, hence of good subjective quality. For the cognitive load, it is reasonable not to wish to complicate the trainee controller's task any further by using an inappropriate

synthetic voice.

10.3.1 Performance of various synthetic voices

First of all, we will examine a number of objective factors for evaluating the quality of the synthetic voice. We begin by making the following distinctions (42):

- **Clarity:** concerns measurements with symbols devoid of meaning (note: in many studies of the English language, clarity is referred to as segmental intelligibility).
- **Intelligibility:** is determined with words or sentences having some meaning, intended to express an idea to be understood.

Without going into detail on the testing methods for clarity, we note that it is simple to find a number of studies that report evaluation results for text-to-speech synthesis systems. The following table shows some of those results (adapted from (43)):

Table XXIIISummary of results of clarity tests, for text-to-speech synthesis systems

Text-to-speech synthesis system	Overall error rate, MRT test, in %
Natural voice	0.53
Dectalk (Paul)	3.25
Dectalk (Betty)	5.72
Prose 3.0	5.72
MITalk-79	7.00
Infovox SA 101	12.50
Berkeley	14.14
TSI-proto 1	17.75
Votrax Type'n'Talk	27.44
Echo	35.56

Note first of all that the synthetic voices evaluated were all poorer than the control (natural) voice. Also, the best performances were by one of the Dectalk system voices (Paul).

These results of clarity and intelligibility evaluations are incomplete, since they do not include the effects of training on different synthetic voices. Accustoming the hearer to a synthetic voice may have a marked effect on its perceived clarity and intelligibility (44).

We were unable to obtain equivalent results for systems employing word or sentence segment concatenation. Nor do there appear to be any comparative studies between text-to-speech synthesis systems and those using word or sentence segment

concatenation.

We mentioned the difference between clarity and intelligibility. The latter takes context into account, i.e. the linguistic and pragmatic elements in the message (45). Familiarity with the task and, insofar as this study is concerned, with a specific phraseology, may considerably affect the intelligibility of a given message. It has been shown that for a given synthetic voice, users familiar with air traffic control phraseology obtained better results than users who were not (46).

Although measurements of the clarity and intelligibility of the synthetic voice are basic considerations, there is also a subjective dimension that takes account of the human listener. Thus we distinguish:

- the natural quality of the voice, i.e. an artificial voice's resemblance to a human voice;
- acceptability, i.e. a human listener's tolerance of an artificial voice.

Since there are many studies evaluating one or more of the above criteria, it is possible to ask which criterion or criteria we should rely on in evaluating and selecting a voice synthesizer. This task is simplified by the links between some of those criteria. It has been shown that there is a strong correlation between the acceptability of a synthetic voice and its perceived clarity (47) (48). Although the results are incomplete (there is no study of how acceptability varies for two voices of the same clarity), it seems likely that the best choice among different voices is the clearest. That is particularly true in that we may wish to use an automated pseudo-pilot to simulate a radio communications channel for training purposes, in which case differences in intelligibility may be increased. To conclude this section, we note that:

- For text-to-speech synthesis systems, the "Paul" voice of the Dectalk system generally obtains the best results.
- There do not seem to be any studies comparing the performances of text-to-speech systems and those using word or sentence segment concatenation.
- There seems to be a positive correlation between the

clarity and acceptability of a voice.

From a perceptual point of view, the impact of a synthetic voice on the user's cognitive load remains to be determined. We will turn to that matter in the next section.

10.3.2 Cognitive load of processing a synthetic voice

There is general agreement that synthetic voices are less clear and less intelligible than human ones (49). But hearing is one thing, and understanding another: thus there is a distinction between perception and comprehension. In addition, the effort of comprehension must not hinder the user in accomplishing his or her principal task.

There are relatively few data on the understanding and use of the information obtained from a synthetic voice. Nevertheless, some results appear to indicate that:

- It is more difficult to remember and use the information received from a synthetic voice (50) (51). It is felt that the perception of synthetic speech calls for greater processing capacity by a user's cognitive resources for encoding the voice.

- More time is required to analyse and use information from a synthetic voice than from a natural one, and that time increases as the quality of the synthetic voice deteriorates (43).

Note that these studies made no attempt to strictly quantify the quality of the synthetic voices used. There are also few comparisons between the comprehension of several synthetic (or artificial) voices. In one case, however, a natural voice, digitized voices (at 9.6 kbits/sec and at 2.4 kbits/sec) and a synthetic voice (Dectalk) were compared (52). No difference was found between the comprehension results for the natural voice and the better-quality digitized voice (9.6 kbits/sec); however, it was found that comprehension results were poorer for the lower-quality digitized voice (2.4 kbits/sec) and the synthetic voice (Dectalk).

As noted in our discussion of intelligibility, practice and training may improve the perception (intelligibility) of the

synthetic voice. However, opinion differs on whether that improvement in perception results in a reduction of the processing capacity needed by the operator. On the basis of the findings in this section, we conclude that:

- The results of comprehension and retention tests are better for the most intelligible voices.

10.4 Effects of artificially generated speech on language behaviour

In this section we will examine the impact of two effects of the use of artificially generated speech on the user's language behaviour:

- a low-level effect--adaptation; and
- a high-level effect--stabilization of the dialogue.

The terms low-level and high-level approximately describe the complexity of the cognitive mechanisms involved.

10.4.1 Adaptation

Adaptation is a well known factor in social psychology. It is recognized that speakers frequently spontaneously adapt their language to that of their interlocutor: we speak of "convergence" when an individual makes his or her language similar to that of the other speaker, and "divergence" when he or she makes it different (22). In general, convergence is associated with a polite and co-operative attitude, while divergence accentuates social or group differences.

Adaptation can play a role at various levels, from sentence construction to the choice and pronunciation of words.

It has also been noted that certain speakers tend to imitate the vocal prompting of a synthetic voice when they begin voice training (providing voice samples) on a speech recognizer (45).

10.4.2 Stabilization of dialogue

We noted in a previous chapter that interaction with a machine can lead to stabilization of dialogue. In fact, one of the tasks of the interface (in this case, the dialogue between the trainee controller and the automated pseudo-pilot) is to provide a representation of the system that the user is dealing with (53). This suggests one of the roles that the synthetic voice may be called upon to play, e.g.:

- To create the perception for the user that he or she is conversing with a system able to understand and interpret his or her instructions, but able to understand a limited language formulated according to relatively strict rules in terms of sentence structure, vocabulary and pronunciation.

We therefore suggest that it may be possible to exploit certain features of the synthetic voice to constantly remind users that they must "watch their language," since they are not dealing with a human. Some researchers are in fact beginning to become interested in this aspect of interfaces using speech generation and recognition (54). For example, the pseudo-pilot's replies could be generated with a high-quality voice, but stressing conformity with official phraseology, more so than the human pseudo-pilot now does. It has been noted that human pseudo-pilots often introduce some variability in their messages. The fact that the messages generated comply strictly with standard phraseology might help standardize the dialogue, by constantly reminding trainees that they are dealing with a machine. There might also be a benefit in terms of the trainees learning the phraseology.

10.5 Selecting a synthetic voice for the automated pseudo-pilot

Our purpose in this section is not to suggest a final and specific choice of an artificial voice for the automated pseudo-pilot, but rather a process that takes account of the factors identified thus far, and can be adapted as necessary as the automated pseudo-pilot prototype is developed.

10.5.1 Requirements and possibilities of existing methods

We want an artificial voice that:

- allows simple modifications of the vocabulary or replies by the automated pseudo-pilot, if necessary;
- possesses high intelligibility, to encourage error-free dialogue, avoid increasing the trainee's work load, and permit the greatest possible acceptability of the system;
- encourages favourable language behaviour by trainees, by constantly reminding them that they are dealing with a system that can perform only in a limited area of natural language;
- makes it possible to generate different voices, to represent different pseudo-pilots with whom a trainee must deal during an exercise;
- can adapt to a stable vocabulary and a closed set of replies by the automated pseudo-pilot.

The most flexible and simplest approach from the point of view of possible modifications to the vocabulary and replies of the pseudo-pilot is text-to-speech synthesis. This approach also has the merit of making it clear to trainees that they are dealing with a machine. Unfortunately, the intelligibility of the best available text-to-speech system (Dectalk, with the "Paul" voice) is not as good as natural speech, and so there are possible drawbacks in terms of comprehension, workload for trainees and acceptability. That is all the more true if the other voices available on the system are used. One last inconvenience is that Dectalk can be used only in English, and so could not be employed for a pseudo-pilot operating in French (if necessary, a text-to-speech synthesizer designed to operate in French could always be employed).

The other approach is word and sentence segment concatenation. Possible modifications to the pseudo-pilot's vocabulary and replies are not as simple to implement, for each new word or sentence segment must be recorded. It is possible to produce as many voices as desired (which nevertheless requires additional storage capacity). This is an important point for adding to the realism of training. It would be possible to give an

appropriate accent to the voices associated with each callsign (e.g., a German accent for Lufthansa, a British accent for British Airways, etc.). In addition, with appropriate division of sentences, it would probably be possible to obtain overall intelligibility (i.e. in terms of each sentence) close to that of a natural voice. The voice obtained lacks prosody, however. Research is now being conducted to improve the prosody of concatenated voices, and it is already possible to obtain concatenated voices that can barely be distinguished from natural voices. This improvement calls for additional work in generating sentences, however.

Note, finally, that this approach is the one chosen by firms handling the speech aspects of pseudo-pilots now under development (ITT, Verbex, Speech Systems Inc.).

10.5.2 Approach proposed

We propose the following approach:

- Generate a vocabulary and a set of responses using word and sentence segment concatenation for a number of voices, with no correction for prosody.

- Test the vocabulary and all the responses on a prototype. Evaluate its acceptability, intelligibility and effect on workload in comparison with a natural voice (through a "Wizard of Oz" simulation), for simulation exercises.

On the basis of those results it will be possible to conclude either that the "straight" concatenated voice is satisfactory, or that the feasibility of a concatenated voice with prosody must be evaluated. In the latter case, the hoped-for improvements and the possible disadvantages (additional processing) will have to be weighed.

11. CONCLUSION

This study first allowed us to identify the main organizations involved in the development of automated pseudo-pilots, in North America and Europe.

We then developed an operating language model suited to various types of simulation exercises used for the training of air traffic controllers in Canada. We proposed an approach allowing us to design a system suited to simple exercises, with all the necessary potential for use in more complex exercises.

Then we presented various recommendations on the use of speech recognition and generation technologies, taking account of the task to be done, the capabilities of the technology and the ergonomics of the dialogue.

Finally, an examination of work done elsewhere and the exercises observed at the Transport Canada Training Institute in Cornwall, as well as the work on modelling the operating language, allow us to conclude that:

- The design of an automated pseudo-pilot for realistic interaction with trainee controllers is possible, at least for simple exercises with novice trainees.

- For interaction between trainees and the automated pseudo-pilot similar to that now in place with human pseudo-pilots, a frame-based language will probably be necessary. This model will probably also have to contain knowledge of the task, the dialogue history and so on.

- It would likely be a good idea to obtain the voice samples required by the recognizer during the phraseology training itself. For example, it may be possible to use the speech recognizer to teach the trainees certain basic aspects of the phraseology.

Nevertheless, aside from the technical feasibility of such a project, we feel that there are other questions to be answered before beginning the development of an automated pseudo-pilot.

First, it is essential to determine the desired impact of a pseudo-pilot on controller training. Is the purpose simply to provide trainees with a system that will respond to their orders, or are we also looking to provide an automated advisor to guide them (such as ATCoach, for example)? If a system including an automated advisor is selected, it will be necessary to evaluate whether the training it provides meets the needs of trainee controllers in the Canadian context.

Second, the operating environment for the system should be defined. Are there to be stand-alone workstations, or stations linked in a network, with hand-offs between sectors?

When these questions have been answered it will be possible to plan the development of an automated pseudo-pilot system that will really meet the training needs of controllers. Then it will be necessary to make a choice: should there be an attempt to build a wholly Canadian system, or should a system designed elsewhere be used? Our observations show that developing an automated pseudo-pilot system is a long-term task. We feel that it would be advantageous to draw on developments elsewhere, perhaps through technology transfer. Adapting those technologies to the Canadian context would probably call for major efforts, but the chances of success would most likely be greater than if a system were to be developed here, starting from zero. In addition, the most attractive features of various systems could then be combined (e.g., a frame-based language system with an automated advisor).

REFERENCES

(1) Ratzer, G. F. G., Low cost, high performance ATC simulation available, The ICAO Bulletin, Sept. 1987, p. 26-28

(2) Swan, P., ATS Inc., Private communication, Montreal, 20 July 1990

(3) Long, G., Federal Aviation Administration, Private communication, during the Speech Tech '90 Conference, 24-26 April 1990, New-York, NY

(4) Long, G., Federal Aviation Administration, Private communication, Montreal, Quebec, July 1990

(5) Grady, M. W., Porter, J. E., Satzer, W. J., Speech understanding in air intercept controller training system design, Technical Report NAVTRAEQUIPCEN 78-C-0044-1, Naval Training Equipment Center, Orlando, FLA, January 1979, 68 p.

(7) Klein, P., Making the skies friendlier, Sun Tech Journal, May\June 1990, p. 117-118

(8) Koble, H., ITT, Private communication, during the Speech Tech '90 Conference, 24-26 April 1990, New-York, NY

(9) Meyn, C., Logicon, Private communication, 12 July 1990, San Diego, CA

(10) Scardina, J. A., Ryberg, P., Gerstenfeld, A., Future ATC automation aids based upon AI technology, Proceedings of the IEEE, Special Issue on Air Traffic Control, vol. 77, no. 11, Nov. 1989, p. 1625-1633

(11) Gerstenfeld, A., Speech recognition integrated with ATC simulation, The ICAO Bulletin, May 1988, p. 22-23

(12) Gerstenfeld, A., Air traffic control training : in full console or PC, The Society for Computer Simulation, Simulators Series, Edited by A. Sharon and M. R. Fakory, vol. 22, no. 2, p. 316-318

(13) Olorenshaw, L., Speech Systems Inc., Private communication, during the Speech Tech '90 Conference, 24-26 April 1990, New-York, NY

(14) Benson, P., Vensko, G., A spoken language understanding system for phraseology training of air traffic controllers, Speech Technology, Media Dimensions, vol. 5, no. 1, Oct/Nov 1989, p. 64-69

(15) Falzon, P., Ergonomie cognitive du dialogue, Presses Universitaires de Grenoble, Grenoble, France, 1989, 175 p.

(16) Néel, F., Matrouf, K., Gauvain, J.L., Mariani, J., Reconnaissance vocale et application en aéronautique, Notes de Recherches, 57e Congrès de l'ACFAS, Montreal, 1989, p. 37-42

(17) Matrouf, K., Néel, F., Gauvain, J.L., Mariani, J., Adaptative syntax representation in an oral task-oriented dialogue for air-traffic controller training, Proceedings of Eurospeech 89, Paris, Sept. 1989, vol. I, p. 187-190

(18) Matrouf, K., Néel, F., Mariani, J., Système de dialogue orienté par la tâche : une application en aéronautique, Journal d'Acoustique, Tome 2, no. 1, mars 1989, p. 85-93

(19) Matrouf, K., Gauvain, J.L., Néel, F., Mariani, J., Adapting probability-transitions in DP matching process for an oral task-oriented dialogue, Proceedings of the 1990 IEEE International Conference on Acoustics, Speech and Signal Processing, Albuquerque, New Mexico, USA, April 3-6, 1990, p. 569-572

(20) Asbury, M. J. A., The application of DVI within an air traffic system, The Official Proceedings of Speech Tech 88, Media Dimensions, Inc., New-York, NY, April 26-28, p. 242-246

(21) Harrison, J. A., Should speech input/output technology be applied to ATC simulators and operational systems?, The ICAO Bulletin, May 1984, p. 29-33

(22) Starr, A., Frankish, C., Human factors aspects of direct voice input in air traffic control operations, CAA contract no. 7D/S/423/6, copy no. 36, 1986, 81 p.

(23) Transport Canada, Air Traffic Services, Air Traffic Control Manual of Operations, up to date on 24 August 1989

(24) Institut de Formation de Transports Canada, Cornwall, Ontario, Midtown ACC, procedures and agreements, DBE-0584, SN 2-2-01-01, June 1989, 31 p.

(25) Plutchak, B., Voice operated status boards for the CATCC, The Official Proceedings of Speech Tech '88, Media Dimensions, Inc., April 26-28, New-York, NY, p. 71-73

(26) Minsky, M., A framework for representing knowledge, from Readings in Knowledge representation, R. J. Brachman and H. J. Levesque (Ed.), Morgan Kaufman Publishers, Inc., Los Altos, CA, 1985, p. 245-262

(27) Abelson, R., Schank, R., Scripts, Plans, Goals and Understanding, Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ, 1977, p. 41

(28) Code d'épellation en radio-téléphonie, in Télécommunications aéronautiques, Volume 2, Annexe 10 à la Convention relative à l'Aviation Civile Internationale, avril 1985, 4e édition

(29) Fiset, J. Y., Descout, R., Robert, J. M., Optimisation de modèles de langage basés sur des schémas, Actes des XVIIIe Journées d'Etudes sur la Parole, Montréal, 28-31 mai 1990, p. 237-241

(30) Holden, J. M., Two helicopter flight tests of voice command, Official Proceedings of Military and Government Speech Tech '89, Media Dimensions Inc., Arlington, VA, Nov. 13-15, 1989, p. 31-35

(31) Pallett, D. S., Performance assessment of automatic speech recognizers, Journal of Research of the National Bureau of Standards, Vol. 90, no. 5, Sept/Oct 1985, p. 371-387

(32) Poock, G. K., Speech recognition research, application and international efforts, Proceedings of the Human Factors Society, 30th Annual Meeting, 1986, p. 1278-1283

(33) Baker, J. D., Pinto, D. F., Optimal and suboptimal training strategies for automatic speech recognition in noise, and the effects of adaptation on performance, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Tokyo, 1986, p. 745-748

(34) Rosenhoover, A., Private communication, during the Speech Tech '90 Conference, 24-26 April 1990, New-York, NY

(35) Luzzati, D., Analyse linguistique d'un corpus de dialogues hommes-machines, Centre de Recherche en Morphosyntaxe du Français Contemporain, Université de Paris III - Sorbonne Nouvelle, Tome 1, 1988, p. 95-97

(36) Calliope, Sunthèse de la parole, which is the chapter XV of La parole et son traitement automatique, J. P. Tubach (Ed), Collection Technique et Scientifique des Télécommunications, Masson, 1989, p. 408-454

(37) Muller, N.J., Voice compression basics : How ADPCM increases channel capacity, Journal of Data & Computer Communications, vol. 2, no. 4, Spring 1990, p. 32-38

(38) Olive, J. P., Nakatani, L. H., Rule-synthesis of speech by work concatenation : A first step, Journal of the Acoustical Society of America, vol. 55, no. 3, March 1974

(39) Young, S. J., Synthesis by rule of prosodic features in Word Concatenation Synthesis, International Journal of Man-Machine Studies, vol. 12, 1980, p. 241-258

(40) Eady, S. J., Dickson, B. C., Urbanczyk, S. C., Claywards, J. A. W., Wynrib, A. G., Pitch assignment rules for speech synthesis by word concatenation, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 3, 1987, p. 1473-1476

(41) Klatt, D. H., Review of text-to-speech conversion for English, Journal of the Acoustical Society of America, vol. 82, no. 3, Sept. 1987, p. 737-793

(42) Calliope, Evaluation de la qualité de la parole codée et synthétique, which is the chapter XVI de La parole et son traitement automatique, J. P. Tubach (Ed.), Collection Technique et Scientifique des Télécommunications, Masson, Paris, 1989, p. 455-488

(43) Streeter, L. A., Applying speech synthesis to user interfaces, from Handbook of Human-Computer Interaction, M. Helander (Ed.), Elsevier Science Publishers B. V., 1988, p. 321-343

(44) Schwab, E. C., Nusbaum, H. C., Pisoni, D. B., Some effects of training on the perception of synthetic speech, Human Factors, vol. 27, no. 4, 1985, p. 395-408

(45) Simpson, C. A., McCauley, M. E., Roland, E. F., Ruth, J. C., Williges, B. H., System design for speech recognition and generation, Human Factors, 1985, vol. 27, no. 2, p. 115-141

(46) Simpson, C. A., Occupational experience with a specific phraseology : Group differences in intelligibility for synthesized and human speech, Journal of the Acoustic Society of America, vol. 58, Supplement no. 1, 1975, DD7

(47) Logan, J. S., Pisoni, D. B., Preference judgements comparing different synthetic voices, in Research on Speech Perception, Progress Report No. 12, Speech Research Laboratory, Indiana University, Bloomington, IN, 1986, p. 262-289

(48) Voiers, W. D., Diagnostic acceptability measure for speech communication systems, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Hartford, Connecticut, USA, May 9-11, 1977, p. 204-207

(49) Pisoni, D. B., Greene, B. G., Logan, J. S., An overview of ten years of research on the perception of synthetic speech, Proceedings of ESCA Tutorial Day and Workshop on Speech Input/Output Assessment and Speech Databases, Noordwijkerhout, The Netherlands, 20-23 Sept. 1989, p. 1.1.1-1.1.4

(50) Luce, P. A., Feustel, T. C., Pisoni, D. B., Capacity demands in short-term memory for synthetic and natural speech, Human Factors, vol. 25, no. 1, 1983, p. 17-32

(51) Waterworth, J. A., The cognitive effort of processing synthetic speech, Eleventh International Symposium on Human Factors in Telecommunications, Sept. 9-13, 1985, CCETT, 35510 Cesson Sevigne, France, A1, p. 1-8

(52) Moody, T. S., Joost, M. G., Synthesized speech, digitized speech, and recorded speech : a comparison of listener comprehension rates, Proceedings of the American Voice I/O Society, 1986, p. 263-280

(53) Nusbaum, H. C., Pisoni, D. B., Human factors issues for the next generation of speech recognition systems, The Official Proceedings of Speech Tech 86, New-York, NY, p. 140-144

(54) Nusbaum, H. C., Pisoni, D. B., Human factors issues for the next generation of speech recognition systems, in Research on Speech Perception, Progress Report No. 12, Speech Research Laboratory, Indiana University, Bloomington, IN, 1986, p. 403-412